

NOTES

“COMMUNITY GUIDELINES”: THE LEGAL IMPLICATIONS OF WORKPLACE CONDITIONS FOR INTERNET CONTENT MODERATORS

*Anna Drootin**

Content moderation is the internet’s not-so-secret, dirty little secret. Content moderators are working around the world, and around the clock, to scrub the internet of horrific content. Most moderators work for low pay and with little or no health care benefits. The content they are exposed to leaves them vulnerable to a number of different mental health issues, including post-traumatic stress disorder. Their work is often hidden from users and is de-emphasized by the technology industry.

This Note explores potential solutions to the labor and employment issues inherent in content moderation work and suggests that there could be a path forward that both empowers and protects workers and leaves technology companies less vulnerable to litigation, bad press, and governmental regulation. An approach that combines corporate and worker-driven social responsibility is the most promising.

INTRODUCTION.....	1198
I. CONTENT MODERATION: A NECESSARY EVIL.....	1200
A. <i>Why Moderate Content?</i>	1201
B. <i>What We Know About the Global Content Moderation Industry</i>	1202
C. <i>Working Conditions</i>	1205
D. <i>Lawsuits</i>	1207
II. PATHS FORWARD.....	1210
A. <i>Status Quo and Litigation</i>	1211
1. <i>The Futility of Continued Litigation and Settlement</i>	1211

* J.D. Candidate, 2022, Fordham University School of Law; B.A., 2013, Rutgers University. I would like to thank Professor James J. Brudney for his invaluable critique, guidance, and encouragement. I would also like to thank Steven Still and the editors and staff of the *Fordham Law Review* for their assistance and dedication. To my husband Tyler, you were a fiancé when I began writing this Note and now you are a husband. Thank you so much for your patience, love, and support.

2. Pitfalls to Inaction	1213
B. <i>Internal Employment Policies and Corporate Social Responsibility</i>	1215
1. Company-Wide Measures	1217
2. Industry-Wide Measures	1219
3. Recommendations from the Experts	1221
C. <i>Regulatory and Legislative Solutions</i>	1223
1. Existing Domestic Legislation and Regulation.....	1224
2. Potential New Domestic Legislation.....	1227
3. International Legislation	1228
a. <i>Legislation Targeting User Safety</i>	1228
b. <i>Business Incentives</i>	1230
c. <i>Transnational Organizations</i>	1231
D. <i>Unionization, Organization, and Worker Mobilization</i> .	1233
1. Transnational Unions	1233
2. Worker-Driven Social Responsibility	1236
3. Current Worker Organization Efforts	1239
III. CREATING “COMMUNITY GUIDELINES” THROUGH CORPORATE AND WORKER-DRIVEN SOCIAL RESPONSIBILITY	1240
CONCLUSION	1243

INTRODUCTION

“If you open a hole on the internet . . . it gets filled with shit.”¹ Whether we use the internet to seek out information, keep in touch with friends, complete tasks for work, or watch an endless stream of TikToks until our eyes bleed, most internet users do not see the dark side of the web. Our relatively pleasant online experience is made possible by the labor of over one hundred thousand content moderators.² Their work ensures that we do not see the “shit.” “Facebook stands for bringing us closer together and building a global community.”³ YouTube’s stated mission is “to give

1. SARAH T. ROBERTS, *BEHIND THE SCREEN: CONTENT MODERATION IN THE SHADOWS OF SOCIAL MEDIA* 165 (2019) (quoting a professional content manager responsible for major online media properties).

2. See Catherine Buni & Soraya Chemaly, *The Secret Rules of the Internet*, VERGE (Apr. 13, 2016), <https://www.theverge.com/2016/4/13/11387934/internet-moderator-history-youtube-facebook-reddit-censorship-free-speech> [<https://perma.cc/GH89-GY9X>] (discussing estimates that the number of workers performing content moderation exceeds 100,000); see also ROBERTS, *supra* note 1, at 23.

3. Mark Zuckerberg, *Building Global Community*, FACEBOOK, <https://www.facebook.com/notes/mark-zuckerberg/building-global-community/10154544292806634/> [<https://perma.cc/T65B-WTQC>] (last visited Sept. 3, 2021). In 2021, Facebook rebranded and changed its company name to “Meta.” Mike Issac, *Facebook Renames Itself Meta*, N.Y. TIMES (Nov. 10, 2021), <https://www.nytimes.com/2021/10/28/>

everyone a voice and show them the world.”⁴ Content moderators are a critical part of making these lofty Silicon Valley mission statements a reality. Their work transforms what would otherwise be an unusable, hostile space into a profitable product.

Scholars, journalists, and concerned citizens have discussed how content moderation impacts society, democracy, voting, and our First Amendment rights at length.⁵ Yet the plight of more than one hundred thousand content moderators working globally is often overlooked amidst the noise about how content moderation aligns with free speech concerns and its effect on users. While large media companies like Facebook tout the fact that they have brought on legal experts and heads of state to analyze complex issues,⁶ they are less vocal about the workers who must apply these policies and directly engage with troubling content all day, every day.

Social media platforms have more power in determining who can speak and be heard around the globe than any judge or head of state.⁷ The consequences of this power were especially evident following the 2016 United States presidential election, “Brexit,” and the exposure of Facebook’s role in stoking violence against the Rohingya community in Myanmar.⁸ The outcry for changes in content moderation has reached such heights as to motivate Facebook to create a global oversight board with final authority over controversial content on the platform.⁹ The board’s members include a former United States federal judge, the first female Danish prime minister, newspaper editors, and human rights activists.¹⁰

While technology companies scramble to address crises driven by user content, journalists have been reporting for years on the poor workplace

technology/facebook-meta-name-change.html [https://perma.cc/67ZP-66Q9]. This Note refers to the company as Facebook for clarity and ease of reference.

4. *About*, YOUTUBE, <https://www.youtube.com/howyoutubeworks/our-mission/> [https://perma.cc/J5FT-PSEC] (last visited Oct. 29, 2021).

5. See generally Buni & Chemaly, *supra* note 2; Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598 (2018).

6. See Nick Clegg, *Welcoming the Oversight Board*, FACEBOOK (May 6, 2020), <https://about.fb.com/news/2020/05/welcoming-the-oversight-board/> [https://perma.cc/U95V-484G].

7. See Miguel Helft, *Facebook Wrestles with Free Speech and Civility*, N.Y. TIMES (Dec. 12, 2010), https://www.nytimes.com/2010/12/13/technology/13facebook.html?_r=0 [https://perma.cc/3LH9-KMJP].

8. See, e.g., Jenny Domino, *Crime as Cognitive Constraint: Facebook’s Role in Myanmar’s Incitement Landscape and the Promise of International Tort Liability*, 52 CASE W. RESV. J. INT’L L. 143, 182–83 (2020); Rebecca Bellan, *Americans Want Transparency in Content Moderation Decisions on Social Media*, FORBES (June 19, 2020, 8:30 AM), <https://www.forbes.com/sites/rebeccabellan/2020/06/19/americans-want-transparency-in-content-moderation-decisions-on-social-media/#6a1539770ae8> [https://perma.cc/VL9D-CZUJ].

9. See Steven Overly & Alexandra S. Levine, *Facebook Announces First 20 Picks for Global Oversight Board*, POLITICO (May 6, 2020, 1:09 PM), <https://www.politico.com/news/2020/05/06/facebook-global-oversight-board-picks-240150> [https://perma.cc/UQ8V-SRKL].

10. See *Expertise from Around the World*, OVERSIGHT BD., <https://oversightboard.com/meet-the-board/> [https://perma.cc/NNU5-YKQJ] (last visited Oct. 29, 2021).

conditions and the mental trauma inflicted on content moderators.¹¹ Workers describe their career in moderation as “three months in hell”¹² and compare themselves to the “sin-eater,” a figure in Welsh and British folklore who was paid to “eat” the sins of a deceased community member so they could enter heaven.¹³ Another worker stated in simple terms that “[t]here was literally nothing enjoyable about the job.”¹⁴

Lawsuits brought by content moderators against large technology companies underscore the need for reform. Workers’ rights, along with their physical and mental health, must be balanced against productivity goals and corporate bottom lines. Companies that employ these workers must acknowledge that content moderators are performing what is arguably the most important work in technology.

Part I of this Note provides a brief background on content moderation, the actual work performed by content moderators, and the lawsuits arising from current working conditions. Part II outlines the potential solutions to the labor and employment issues content moderators face, some of which are currently being implemented, albeit with a lack of uniformity and consistency. Part III recommends a workable solution that benefits both technology companies and content moderators.

I. CONTENT MODERATION: A NECESSARY EVIL

The plight of content moderators as a workforce cannot be addressed without an understanding of how and why content is moderated. Working conditions exposed through journalism and scholarship, as well as in settled and ongoing lawsuits, paint a clear picture that the industry must change.

Part I.A provides a brief history of content moderation and explains why companies moderate. Part I.B presents an overview of the global content moderation industry as a whole, including the number of workers, types of content moderation positions, and where this work is performed. Part I.C describes content moderators’ working conditions. Part I.D discusses legal action taken by content moderators against their employers.

11. See Casey Newton, *The Trauma Floor*, VERGE (Feb. 25, 2019, 8:00 AM), <https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona> [<https://perma.cc/NX76-DBTC>]; Brad Stone, *Policing the Web’s Lurid Precincts*, N.Y. TIMES (July 18, 2010), <https://www.nytimes.com/2010/07/19/technology/19screen.html> [<https://perma.cc/MS2P-GQN8>].

12. Burcu Gültekin Punsmann, *Three Months in Hell*, SÜDDEUTSCHE ZEITUNG MAGAZIN (Jan. 6, 2018), <https://sz-magazin.sueddeutsche.de/internet/three-months-in-hell-84381> [<https://perma.cc/GXT4-U7RE>].

13. See ROBERTS, *supra* note 1, at 165.

14. Olivia Solon, *Underpaid and Overburdened: The Life of a Facebook Moderator*, GUARDIAN (May 25, 2017, 5:00 AM), <https://www.theguardian.com/news/2017/may/25/facebook-moderator-underpaid-overburdened-extreme-content> [<https://perma.cc/5FZW-DK3Q>].

A. Why Moderate Content?

In its early days, the internet was a space with little to no censorship.¹⁵ Moderation occurred, but it was often performed by volunteers enforcing standards based on local rules of engagement around community norms and user behavior.¹⁶ As the internet became a global commodity, content moderation became a vital part of the business of the internet.¹⁷

The development of content moderation into a technology sector of its own was not swift. Facebook, for example, did not form a content moderation team until November 2009, five years after the site went live.¹⁸ Early moderators at Facebook made removal decisions based on a one-page document.¹⁹ They deleted posts that “ma[ke] you feel bad in your stomach.”²⁰ One page eventually grew into a 15,000-word moderation policy, which is now amended frequently.²¹

Complicated moderation schemes did not grow out of government regulation or a desire to limit liability. Section 230 of the Communications Decency Act of 1996 shields internet platforms from liability for user-generated content.²² The purpose of Section 230 was twofold: (1) to encourage platforms to act as good Samaritans taking an active role in removing offensive content and (2) to avoid the free speech issues surrounding platforms’ collateral censorship of users’ speech.²³ The structure of the statute gives platforms broad freedom to choose which values they want to protect or to protect no values at all in their moderation schemes.²⁴

15. See ROBERTS, *supra* note 1, at 5.

16. See *id.*

17. See *id.* at 6.

18. See Klonick, *supra* note 5, at 1620.

19. See *id.* at 1631.

20. See PAUL M. BARRETT, WHO MODERATES THE SOCIAL MEDIA GIANTS?: A CALL TO END OUTSOURCING 7 (2020), https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/5ed9854bf618c710cb55be98/1591313740497/NYU+Content+Moderation+Report_June+8+2020.pdf [<https://perma.cc/VT6G-UVJW>].

21. See *id.* at 7. In comparison, Pinterest’s public-facing community guidelines are around 3000 words. See *Community Guidelines*, PINTEREST, <https://policy.pinterest.com/en/community-guidelines> [<https://perma.cc/S4ZS-QNQU>] (last visited Oct. 29, 2021). Twitter’s rules page contains links to more than seventeen different policies on topics such as hateful conduct, nonconsensual nudity, and impersonation. See *Rules and Policies*, TWITTER, <https://help.twitter.com/en/rules-and-policies#twitter-rules> [<https://perma.cc/VP6T-YHHM>] (last visited Oct. 29, 2021).

22. See generally 47 U.S.C. § 230.

23. See Klonick, *supra* note 5, at 1602; see also *Zeran v. Am. Online, Inc.*, 129 F.3d 327, 330–31 (4th Cir. 1997) (discussing how the statute’s purpose is evident from the plain language enacted by Congress, particularly the findings in subsection (a)).

24. See Klonick, *supra* note 5, at 1617. The Act clearly states that technology companies flourish with minimal government regulation and that companies will not be liable as a publisher of user content, leaving companies free to handle user content however they see fit. See generally, 47 U.S.C. § 230. Compare *Community Standards*, FACEBOOK, <https://www.facebook.com/communitystandards/introduction/> [<https://perma.cc/588Z-6GAZ>] (last visited Oct. 1, 2021), with Ali Breland, *Twitter to Implement Changes Meant to*

Under Section 230, platforms do not have to moderate but many do so anyway.²⁵ The transformation of content moderation from a few Silicon Valley workers using a sheet of paper to a multinational, multitiered technology sector was spurred partially by a sense of corporate social responsibility and largely by economic concerns.²⁶ Creating safe spaces for users aligns with the Silicon Valley ethos of sharing, community, and connection. However, the bottom line is that if these spaces are not hospitable to users, engagement will decrease, and advertising revenue will decrease accordingly.²⁷

Content moderation as we know it today is a hybrid of algorithmic and human analysis.²⁸ Algorithms can catch things like child sexual exploitation material and nudity.²⁹ PhotoDNA, for example, is an algorithm that converts the known universe of child sexual abuse content into grayscale, overlays a grid, and assigns a numerical value to each square creating a signature that remains, even when images are altered.³⁰ Technology like PhotoDNA can lighten the load in certain content categories, but the companies themselves acknowledge that algorithms can filter out too much.³¹ As a result, almost all user-generated content published online is reviewed reactively, through flagging by other users and subsequent review by human content moderators using internal guidelines.³² On Facebook alone, more than three million items are reported daily and flagged for removal.³³ A global, 24/7 workforce is required to perform moderation on such a scale.

B. What We Know About the Global Content Moderation Industry

Professor Sarah Roberts outlines a taxonomy of online content moderator labor arrangements in her book *Behind the Screen*.³⁴ Some moderators work on-site or in-house for a company that requires content screening.³⁵ Others work for boutique firms that specialize in online brand management and perform content moderation for other companies.³⁶ The call center arrangement has arguably attracted the most media attention and likely

Crack Down on Trolls, HILL (May 15, 2018, 2:05 PM), <https://thehill.com/policy/technology/387790-twitter-to-implement-new-changes-in-troll-crackdown> [<https://perma.cc/8VYJ-8ALP>] (discussing Twitter's reputation as a "cesspool of negativity" prior to taking steps to moderate content and accounts).

25. See generally 47 U.S.C. § 230.

26. See Klonick, *supra* note 5, at 1616–18.

27. See *id.* at 1627.

28. See BARRETT, *supra* note 20, at 4.

29. See *id.* at 3.

30. See Klonick, *supra* note 5, at 1636–37.

31. See BARRETT, *supra* note 20, at 4.

32. See Klonick, *supra* note 5, at 1638.

33. See BARRETT, *supra* note 20, at 2.

34. See ROBERTS, *supra* note 1, at 40–43.

35. See *id.* at 41.

36. See *id.* at 41–42.

represents the employment structure for the majority of content moderators.³⁷ Content moderators working at call centers are located in large-scale operations centers with the technological infrastructure to handle multiple contracts for different companies and provide 24/7 services.³⁸ In the United States, third-party vendors such as Accenture, Competence Call Center, CPL Resources, Genpact, and Majorel run these operations.³⁹ Cognizant, the vendor whose workplace conditions were exposed in Casey Newton's 2019 article "The Trauma Floor," has since ceased content moderation operations.⁴⁰ The call center outsourcing model saves companies money and enables them to tap into workers across the globe who speak multiple languages and can work around the clock.⁴¹ Finally, content moderators can perform their work using microlabor platforms, an arrangement that is even more disjointed and disconnected from the parent companies than the call center model.⁴² Under the microlabor platform arrangement, content moderators work remotely worldwide, performing work on a per-task basis using platforms like Amazon Mechanical Turk.⁴³

Some industry experts estimate that there are one hundred thousand workers performing content moderation throughout the world.⁴⁴ Others believe the number of workers is even higher.⁴⁵ Recent data posits that there are 15,000 moderators working for Facebook, 10,000 moderating Google and YouTube products, and 1500 working for Twitter.⁴⁶ This puts the total at over 25,000 for three U.S. platforms alone. China's most popular news application, Jinri Toutiao, expanded its moderation team to 10,000

37. See Newton, *supra* note 11.

38. See ROBERTS, *supra* note 1, at 42.

39. See BARRETT, *supra* note 20, at 12–13.

40. See Newton, *supra* note 11; see also Casey Newton, *Why a Top Content Moderation Company Quit the Business Instead of Fixing Its Problems*, VERGE (Nov. 1, 2019, 6:00 AM), <https://www.theverge.com/interface/2019/11/1/20941952/cognizant-content-moderation-restructuring-facebook-twitter-google> [<https://perma.cc/T4ZW-KHU3>].

41. See BARRETT, *supra* note 20, at 18.

42. See ROBERTS, *supra* note 1, at 43.

43. See *id.* at 40, 43.

44. See ROBERTS, *supra* note 1, at 23; Buni & Chemaly, *supra* note 2; Adrian Chen, *The Laborers Who Keep Dick Pics and Beheadings Out of Your Facebook Feed*, WIRED (Oct. 23, 2014, 6:30 AM), <https://www.wired.com/2014/10/content-moderation> [<https://perma.cc/7LKL-XLDL>]; *Content Moderation: The Future Is Bionic*, ACCENTURE (2017), https://www.accenture.com/cz-en/_acnmedia/PDF-47/Accenture-Webscale-New-Content-Moderation-POV.pdf [<https://perma.cc/RP2Q-W42W>]. There is no definitive source for the number of moderators, partly due to the platforms' reluctance to disclose how much they are outsourcing. Email from Paul M. Barrett, Deputy Dir., N.Y.U. Stern Ctr. for Bus. and Hum. Rts. (Nov. 24, 2020, 3:39 PM) (on file with author).

45. See Buni & Chemaly, *supra* note 2; see also JANINE BERG ET AL., DIGITAL LABOUR PLATFORMS AND THE FUTURE OF WORK 13 (2018), https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_645337.pdf [<https://perma.cc/3TFS-DANM>] (estimating that 150,000 workers are moderating content).

46. See BARRETT, *supra* note 20, at 2; cf. Casey Newton, *Bodies in Seats*, VERGE (June 19, 2019, 8:00 AM), <https://www.theverge.com/2019/6/19/18681845/facebook-moderator-interviews-video-trauma-ptsd-cognizant-tampa> [<https://perma.cc/J84D-BUEV>] (estimating that 30,000 employees were working on safety and security at Facebook in 2019).

workers.⁴⁷ Chinese video streaming application Kuaishou may employ up to 5000 moderators.⁴⁸ Data is not easy to come by. Facebook, for example, refuses to disclose the exact number and location of moderating hubs that it utilizes around the world.⁴⁹ In addition, it should be noted that there are content moderators working for governments.⁵⁰

In the United States, litigation and journalism have exposed in-house moderators and contractors working in California,⁵¹ Washington,⁵² Florida,⁵³ Arizona,⁵⁴ and Texas,⁵⁵ but content moderation work in some form is likely occurring in all fifty states. Companies like Aureon (formerly Caleris) are in the business of “domestic outsourcing” and offer social media moderation services on a company-based level.⁵⁶ Aureon is based in Iowa, and its employees moderate the social media accounts of individual brands.⁵⁷ Overseas, content moderation and customer service call centers have become a national industry in the Philippines.⁵⁸ Facebook has moderators working in Germany, India, Ireland, Kenya, Latvia, the Philippines, Portugal, and Spain.⁵⁹ Germany has become a particularly important moderation hub in the wake of its “NetzDG” Network Enforcement Act, which requires social media companies to remove hate speech and violence within set time limits.⁶⁰

47. See Zhang Yu & Xie Wenting, *China's Huge Pool of Web Moderators Required to Have an Eagle Eye for Dangerous Content*, GLOBAL TIMES (Apr. 16, 2018, 6:33 PM), <https://www.globaltimes.cn/content/1098173.shtml> [<https://perma.cc/57X8-V62G>].

48. See *id.*

49. See Nick Hopkins, *Facebook Moderators: A Quick Guide to Their Job and Its Challenges*, GUARDIAN (May 21, 2017, 1:00 PM), <https://www.theguardian.com/news/2017/may/21/facebook-moderators-quick-guide-job-challenges> [<https://perma.cc/VTN8-ULAQ>].

50. See Kyle Langvardt, *Regulating Online Content Moderation*, 106 GEO. L.J. 1353, 1362 (2018) (discussing the roughly one hundred thousand workers policing the internet in China to remove offensive content as quickly as possible). This Note focuses on commercial moderation of horrific content, not content moderation undertaken by governments to control political or cultural speech.

51. See Complaint for Declaratory and Injunctive Relief ¶ 3, *Scola v. Facebook, Inc.*, No. 18-CIV-05135 (Cal. Super. Ct. Sep. 21, 2018); ROBERTS, *supra* note 1, at 74.

52. See Complaint for Damages ¶¶ 2.1–2.2, *Soto v. Microsoft Corp.*, No. 16-2-31049-4 (Wash. Super. Ct. Dec. 30, 2016).

53. See Plaintiffs' Amended Class Action Complaint and Demand for Jury Trial ¶ 2, *Garrett v. Facebook, Inc.*, No. 8:20-CV-00585 (M.D. Fla. Mar. 12, 2020).

54. See *id.*

55. See Casey Newton, *YouTube Moderators Are Being Forced to Sign a Statement Acknowledging the Job Can Give Them PTSD*, VERGE (Jan. 24, 2020, 10:15 AM), <https://www.theverge.com/2020/1/24/21075830/youtube-moderators-ptsd-accenture-statement-lawsuits-mental-health> [<https://perma.cc/JXL9-JM82>].

56. See ROBERTS, *supra* note 1, at 62–64.

57. See *Social Media Moderation*, AUREON, <https://www.aureon.com/services/customer-service/customer-care/social-media/> [<https://perma.cc/3464-ANR4>] (last visited Oct. 29, 2021).

58. See ROBERTS, *supra* note 1, at 183; IM SCHATTEN DER NETZWELT (THE CLEANERS) (Gebueder Beetz Filmproduktion 2018) (documenting the experience of moderators working in the Philippines).

59. See BARRETT, *supra* note 20, at 3.

60. See *infra* Part II.C.3.a.

Much like the lack of consensus on the number of content moderators working globally, content moderators' precise locations are difficult to pin down. This is both intentional on the part of technology companies and inherent to the nature of the work. The moderators' quick turnover rates and the nondisclosure agreements their companies require them to sign make the moderators difficult to study.⁶¹ Facebook consistently denies requests to visit moderation sites, claiming they cannot disclose their locations to protect the moderators who work there from retaliation by angry users whose content has been removed.⁶² Facebook has not publicly identified specific instances of retaliation or indicated what exactly angry users have said or threatened to do in response to the removal of their content.

Despite the large number of content moderators working across the globe, content moderation is relegated to a second-class function.⁶³ It does not fit into Silicon Valley's engineering and marketing culture and tends to be a source of bad press.⁶⁴ The idea that moderation is a necessary evil, but not part of the essential business of technology, has led to low wages and poor working conditions in the industry.

C. Working Conditions

The typical day for a worker in Cognizant's Facebook content moderation center in Phoenix involved locking away all personal belongings and phones in a locker.⁶⁵ Reviewers then logged on to propriety software known as the Single Review Tool.⁶⁶ When they were ready to work, they clicked "Resume Reviewing" and posts appeared in their queue in no particular order.⁶⁷ Some posts were violent, some depicted nudity or sexual activity, and others included bullying, hate speech, or racism.⁶⁸ "Wellness" break times, bathroom breaks, time for prayer, and lunches were heavily monitored.⁶⁹ Silicon Valley workers interviewed by Professor Roberts in *Behind the Screen* described a process by which content that violated different guidelines was sorted into different queues.⁷⁰ Moderators reviewed batches of videos within a chosen queue by using thumbnails or by watching the full video.⁷¹ These workers estimated that, in total, they reviewed between 1500 and 2000 videos per day.⁷² Facebook contract workers in Ireland reported analyzing between 600 and 800 pieces of content over the course of a typical

61. See ROBERTS, *supra* note 1, at 25–26.

62. See BARRETT, *supra* note 20, at 5; Newton, *supra* note 11.

63. See *id.* at 8.

64. See *id.*

65. See Newton, *supra* note 11.

66. See *id.*

67. See *id.*

68. See *id.*

69. See *id.*

70. See ROBERTS, *supra* note 1, at 88.

71. See *id.*

72. See *id.*

eight-hour shift and up to one thousand pieces on a particularly busy day.⁷³ Facebook workers, in particular, described a highly subjective quality assurance process designed to ensure moderators were following the company's moderation guidelines.⁷⁴ If a quality assurance reviewer came to a different decision on a piece of content, it would reduce the individual's accuracy scores.⁷⁵ This process created constant tension in the workplace.⁷⁶

Throughout their day, workers encounter images, video, and text that depict adult nudity and sexual activity; violent and graphic content, including harm to both people and animals; content from dangerous organizations, like terror groups or organized hate groups; hate speech; drugs and firearms; child nudity and sexual exploitation of children; bullying and harassment; and suicide and self-injury.⁷⁷ Silicon Valley workers interviewed by Professor Roberts indicated that footage from war-torn areas, graphic depictions of sexual abuse involving children, and cases of self-harm threats caused them the most trauma.⁷⁸ Workers in Facebook's Phoenix moderation center expressed that they began to embrace fringe viewpoints and conspiracy theories after being repeatedly exposed to them.⁷⁹ Some of the worst content comes back to haunt moderators when it is reuploaded hundreds or even thousands of times by users.⁸⁰

Constant exposure to horrifying content takes a toll on moderators. Effects include insomnia, nightmares, unwanted memories of troubling images, anxiety, depression, and emotional detachment.⁸¹ One worker claimed he did not see an impact outside of the workplace, but he gained weight, increased his alcohol consumption, and struggled with thoughts about specific images or videos after work.⁸² Workers at Facebook's Phoenix center used drugs and alcohol and engaged in sexual activities in the workplace.⁸³

One moderation center contractor has openly acknowledged the risks to content moderation employees. Accenture asked employees at Facebook and YouTube sites in Texas and in Europe to sign a two-page form attesting to

73. See BARRETT, *supra* note 20, at 13; Newton, *supra* note 11.

74. See BARRETT, *supra* note 20, at 13; Newton, *supra* note 11. At the Cognizant centers, Facebook selected fifty to sixty random decisions to audit, which would be reviewed by a quality assurance worker at the vendor. *See id.* A subset of those quality assurance decisions would be audited by full-time Facebook employees. *See id.*

75. See BARRETT, *supra* note 20, at 17; Newton, *supra* note 11.

76. See BARRETT, *supra* note 20, at 17; Newton, *supra* note 11 (discussing an employee's opinion that the scores were "fake" because they measured agreement between quality assurance reviewers and moderators, which is an unhelpful standard when the work is inherently subjective). For instance, whether or not a quality assurance reviewer would regard an overly lenient decision as equally inaccurate as an overly restrictive decision is unclear.

77. See BARRETT, *supra* note 20, at 10–11.

78. See ROBERTS, *supra* note 1, at 105–06.

79. See Newton, *supra* note 11.

80. See BARRETT, *supra* note 20, at 13.

81. *See id.* at 14.

82. See ROBERTS, *supra* note 1, at 112.

83. See Newton, *supra* note 11.

the dangers and the risk that the job could lead to post-traumatic stress disorder (PTSD).⁸⁴ The form asked workers to take advantage of workplace wellness programs but acknowledged that the “wellness coach” provided by Accenture was not a medical doctor and could not diagnose or treat mental health disorders.⁸⁵

Workers face barriers in seeking the help they need for these side effects, even when the companies try to provide it through “wellness” initiatives or other programs.⁸⁶ Workers are hesitant to admit that content is affecting them because seeking help might be equated to not having the skills to master the job.⁸⁷ Content moderators express that the content they view is so traumatic they are hesitant to burden even trained counselors by describing it, let alone to lean on family or friends for support.⁸⁸ Workers allege that structures intentionally put in place by technology companies create hurdles to seeking counseling.⁸⁹ Many content moderators are required to sign nondisclosure agreements, which employers claim protect user privacy.⁹⁰ However, these agreements cut employees off from confiding in others and keep the industry shrouded in secrecy.⁹¹ Evidence suggests that moderators rarely last more than one or two years working under these conditions, and in some cases they are required to leave at a set time.⁹²

The call center model creates layers between the technology companies and moderators, allowing for plausible deniability when issues arise stemming from the work.⁹³ Overseas workers, in particular, are hamstrung by geography, jurisdiction, and bureaucracy in raising workplace complaints.⁹⁴ Faced with these workplace issues, many moderators have looked to the courts for redress.

D. Lawsuits

Content moderators in the United States and abroad have brought legal complaints in court alleging severe mental injury and trauma from their working conditions. This section discusses three cases filed in U.S. state courts and one case pending in an Irish court.

84. See Madhumita Murgia, *Facebook Content Moderators Required to Sign PTSD Form*, FIN. TIMES (Jan. 26, 2020), <https://www.ft.com/content/98aad2f0-3ec9-11ea-a01a-bae547046735> [<https://perma.cc/YR4A-HWGB>]; Newton, *supra* note 55.

85. See Murgia, *supra* note 84.

86. See ROBERTS, *supra* note 1, at 113.

87. See *id.*

88. See *id.* at 118–19.

89. See Plaintiffs’ Amended Class Action Complaint and Demand for Jury Trial, *supra* note 53, ¶ 67; ROBERTS, *supra* note 1, at 35–36.

90. See Newton, *supra* note 11.

91. See *id.*

92. See ROBERTS, *supra* note 1, at 25–26, 74, 124–25; see also *infra* note 190 and accompanying text.

93. See ROBERTS, *supra* note 1, at 194–95.

94. See *id.* at 195.

In 2016, Henry Soto and Greg Blauert, together with their spouses, filed suit against Microsoft in Washington Superior Court.⁹⁵ Soto and Blauert were members of Microsoft's Online Safety Team.⁹⁶ The Online Safety Team reviewed photos and videos depicting brutality, murder, sexual assault, and death.⁹⁷ Soto and Blauert alleged that Microsoft failed to warn them about the toxic effects of the content.⁹⁸ Soto experienced sleep disturbance, nightmares, an internal video screen in his head where he could see disturbing images, irritability, anticipatory anxiety, distraction, and auditory hallucinations.⁹⁹ He began to have panic attacks in public, disassociate, and experience depression and hallucinations.¹⁰⁰ He was unable to be around computers or young children, including his own son.¹⁰¹ Blauert suffered a physical and mental breakdown in 2013.¹⁰² He experienced psychomotor retardation, intractable crying, insomnia, and anxiety.¹⁰³ When the complaint was filed, he remained in treatment for PTSD.¹⁰⁴ Soto attempted to bring a workers' compensation claim but was denied.¹⁰⁵ Their complaint alleged negligence, negligent infliction of emotional distress, violations of the Washington Disability Discrimination Act and the Washington Consumer Protection Act, and loss of consortium.¹⁰⁶ Washington Superior Court records indicate the case settled in 2019.¹⁰⁷

Selena Scola, a Silicon Valley-based moderator, sued Facebook and her contractor-employer, PRO Unlimited, Inc. in 2018 in California Superior Court.¹⁰⁸ Scola alleged that she developed significant psychological trauma and PTSD as a result of her exposure to disturbing images throughout the course of her work.¹⁰⁹ Scola's complaint alleged that Facebook was not providing its content moderators with sufficient training and was not implementing safety standards. The complaint also accused the company of negligence and violations of California's Unfair Competition Law.¹¹⁰ The complaint was styled as a class action on behalf of all California citizens who

95. *See generally* Complaint for Damages, *supra* note 52.

96. *See id.* ¶¶ 3.4, 3.33.

97. *See id.* ¶ 3.16.

98. *See id.* ¶ 3.5.

99. *See id.* ¶ 3.24.

100. *See id.* ¶ 3.26.

101. *See id.*

102. *See id.* ¶ 3.36.

103. *See id.* ¶ 3.37.

104. *See id.*

105. *See id.* ¶¶ 3.28–3.29. Washington state law authorizes a narrow set of claims arising out of mental stress. For a detailed discussion on the potential for workers' compensation claims to address content moderators' workplace injuries, see *infra* Part II.C.1.

106. *See generally* Complaint for Damages, *supra* note 52.

107. *See* Order for Dismissal with Prejudice, *Soto v. Microsoft Corp.*, No. 16-2-31049-4 (Wash. Super. Ct. Feb. 15, 2019).

108. *See generally* Complaint for Declaratory and Injunctive Relief, *Scola v. Facebook, Inc.*, No. 18-CIV-05135 (Cal. Super. Ct. Sept. 21, 2018).

109. *See id.* ¶ 4.

110. *See generally id.*

performed content moderation for Facebook within three years.¹¹¹ In 2020, Scola filed an amended complaint adding class representatives representing workers in Arizona, Florida, and Texas.¹¹²

Facebook's answer highlights the benefits large technology companies reap as a result of the call center model for content moderation.¹¹³ In addition to the affirmative defenses that workers' compensation was the exclusive remedy for Scola's claims and that Scola assumed the risk inherent in content moderation, Facebook argued that Scola's status as an independent contractor, the negligence of third parties, and the lack of an affirmative duty shielded the company from liability.¹¹⁴ PRO Unlimited similarly argued that Scola's exclusive remedy lied in workers' compensation.¹¹⁵

On July 14, 2021, the court granted Scola's motion for final approval of settlement.¹¹⁶ The parties agreed to a \$52 million class action settlement, which provides a base payment of one thousand dollars to content moderators who performed work in Arizona, California, Florida, or Texas, plus further payments up to \$50,000 based on a qualified diagnosis.¹¹⁷ In addition to monetary payments, Facebook agreed that it will require its U.S. vendors to retain licensed and certified clinicians who are experienced in mental health counseling, are familiar with symptoms of PTSD, and will be available during every shift to speak with workers.¹¹⁸ Facebook committed to holding group wellness sessions monthly and making weekly one-on-one coaching or wellness sessions available.¹¹⁹ The settlement stipulates that the company and contractors will provide clear guidelines for how and when moderators can remove themselves from specific tasks involving graphic and objectionable content.¹²⁰ Facebook and its contractors must provide information to moderators about these new policies, including a telephone hotline where compliance concerns can be reported directly to Facebook.¹²¹

111. *See id.* ¶¶ 56–59.

112. *See generally* Second Amended Complaint and Demand for Jury Trial, *Scola v. Facebook, Inc.*, No. 18-CIV-05135 (Cal. Super. Ct. June 30, 2020).

113. *See generally* Defendant Facebook, Inc.'s Answer to Complaint for Declaratory and Injunctive Relief, *Scola v. Facebook, Inc.*, No. 18-CIV-05135 (Cal. Super. Ct. Nov. 5, 2018).

114. *See id.* at 2–3, 6–7, 9.

115. *See generally* Defendant PRO Unlimited, Inc.'s Notice of Demurrer and Demurrer; Memorandum of Points and Authorities, *Scola v. Facebook, Inc.*, No. 18-CIV-05135 (Cal. Super. Ct. Nov. 8, 2018). When Scola amended her complaint to join the claims of two other content moderators who worked for PRO Unlimited and Accenture in California, she dropped PRO Unlimited as a defendant. *See generally* Amended Complaint and Demand for Jury Trial, *Scola v. Facebook, Inc.*, No. 18-CIV-05135 (Cal. Super. Ct. Mar. 1, 2019).

116. *See generally* Order and Judgment Granting Plaintiff's Renewed Motion for Final Approval of Settlement, *Scola v. Facebook, Inc.*, No. 18-CIV-05135 (Cal. Super. Ct. July 14, 2021).

117. *See* Proposed Order Granting Preliminary Approval of Settlement, Ex. 2 at 4–5, 8, *Scola v. Facebook, Inc.*, No. 18-CIV-05135 (Cal. Super. Ct. Aug. 12, 2020).

118. *See id.* at 5.

119. *See id.*

120. *See id.*

121. *See id.* at 6.

On February 5, 2020, Debryna Garrett and Clifford Jeury filed suit against Facebook and Cognizant Business Services Corporation personally and on behalf of all Florida citizens moderating content for Facebook.¹²² The complaint was later amended, removing Jeury as a named plaintiff and expanding the class to include Arizona citizens who performed content moderation work for Facebook as Cognizant employees.¹²³ The amended complaint alleged deliberate concealment or misrepresentation of a known danger, negligence, negligent provision of unsafe equipment, and violations of the Florida Deceptive and Unfair Trade Practices Act.¹²⁴ Individual plaintiffs' alleged injuries included diagnoses of PTSD and related mental health impairments.¹²⁵ The amended complaint also suggested that an employee died during his shift as a result of a heart attack brought on by viewing content at his desk.¹²⁶ The case was removed to federal court, and it is currently pending.¹²⁷

Facebook is also involved in litigation concerning content moderation overseas. Former moderator Chris Gray is suing Facebook Ireland and contractor CPL Solutions in Irish High Court alleging psychological injuries from repeated and unrelenting exposure to extremely disturbing, graphic, and violent content.¹²⁸ Gray's case and the cases of other Facebook Ireland moderators are supported by Foxglove, a nonprofit team of lawyers, technology experts, and communications experts seeking to hold governments and large companies accountable for their abuse of technology.¹²⁹ In order to explore the possibility of bringing further legal action, Foxglove has put a call out to content moderators around the world to reach out and share their experiences.¹³⁰

II. PATHS FORWARD

How do we solve what an industry insider described as a “one-billion dollar problem?”¹³¹ Part II.A of this Note explores the likely outcome if large technology companies maintain the status quo. Part II.B discusses the

122. See generally Class Action Complaint and Demand for Jury Trial, Garrett v. Facebook, Inc., No. 20-CA-001146 (Fla. Cir. Ct. Feb. 5, 2020).

123. See generally Amended Class Action Complaint and Demand for Jury Trial, Garrett v. Facebook, Inc., No. 20-CA-001146 (Fla. Cir. Ct. Mar. 6, 2020).

124. See generally *id.*

125. See *id.* ¶¶ 78, 85, 88.

126. See *id.* ¶ 75.

127. See Complaint and Notice of Removal, Garrett v. Facebook, Inc., No. 8:20-cv-00585 (M.D. Fla. Mar. 12, 2020).

128. See Simon Carswell & Jennifer O'Connell, *Facebook Facing Up to 12 Lawsuits over 'Disturbing Content,'* IRISH TIMES (Dec. 5, 2019, 2:03 PM), <https://www.irishtimes.com/news/crime-and-law/facebook-facing-up-to-12-lawsuits-over-disturbing-content-1.4104871> [<https://perma.cc/J54M-APRD>].

129. See *Who We Are*, FOXGLOVE, <https://www.foxglove.org.uk/about> [<https://perma.cc/8VCR-YB6T>] (last visited Oct. 29, 2021).

130. See Cori Crider, *Why I'm So Keen to Talk to Facebook Content Moderators*, FOXGLOVE (July 22, 2020), <https://www.foxglove.org.uk/2020/07/22/why-im-so-keen-to-talk-to-facebook-content-moderators/> [<https://perma.cc/CC7E-XV9W>].

131. See ROBERTS, *supra* note 1, at 206.

potential for internal reform. Part II.C explores possible domestic and international legislative and regulatory solutions. Part II.D discusses the potential benefits of unionization or worker organization, locally and globally.

A. *Status Quo and Litigation*

Technology companies can continue on the path they are currently on and address the physical and emotional toll on content moderators only when confronted via litigation or through haphazard and inconsistently applied employee wellness programs. A head-in-the-sand approach that ignores the issues inherent to content moderation may work in the short term, but as an understanding of how content moderators are affected by their work deepens, the issues plaguing other industries with similar workplace safety challenges highlight what could be in store. The following section discusses the National Football League and the international garment industry as useful comparators.

1. The Futility of Continued Litigation and Settlement

In January 2012, the more than 240 lawsuits filed by current and former players against the National Football League (NFL) were consolidated into what later became known as the “Concussion Litigation.”¹³² Plaintiffs alleged the NFL owed them a variety of duties including a duty to inform or disclose the risks associated with brain injuries in football, a duty to protect players, and a duty to competently study the risks of brain injuries in football.¹³³ The Eastern District of Pennsylvania did not have to rule on whether the NFL owed these duties as a matter of law because the parties reached a settlement in 2014.¹³⁴ The settlement compensates players who retired from playing NFL football before the preliminary approval of the class settlement on July 7, 2014.¹³⁵ The agreement created an uncapped monetary award fund that entitles players to awards based on diagnosis, age, and the number of seasons played.¹³⁶ The class consisted of over 20,000 retired players,¹³⁷ and the fund approved more than \$500 million in claims

132. See Christopher R. Deubert et al., *Protecting and Promoting the Health of NFL Players: Legal and Ethical Analysis and Recommendations*, 7 HARV. J. SPORTS & ENT. L. 1, 184 (2016).

133. See *id.* at 185.

134. *In re Nat'l Football League Players' Concussion Injury Litig.*, 301 F.R.D. 191, 195 (E.D. Pa. 2014).

135. See *id.* at 204.

136. See *id.* at 196–97.

137. See Deubert et al., *supra* note 132, at 23.

in its first two years.¹³⁸ Attorneys for the players estimate the settlement payout could reach \$1.4 billion.¹³⁹

The NFL Head, Neck, and Spine Committee developed the NFL Game Day Concussion Diagnosis and Management Protocol (“Concussion Protocol”) in 2011.¹⁴⁰ The Football Players Health Study at Harvard University interviewed current players, former players, and contract advisors about their reaction to the Concussion Protocol.¹⁴¹ Responses indicated a mixed bag in terms of perceptions of effectiveness.¹⁴² Some players expressed the sentiment that the protocol was more about public appearance and litigation avoidance than improving player safety.¹⁴³

The settlement itself, the Concussion Protocol, and efforts by the National Football League Players Association (NFLPA) to educate players in the years following the settlement raise the question of whether players expressly assume the risk of brain injury when they take the field.¹⁴⁴ The NFL is protected against future lawsuits due to its collective bargaining agreement, which it negotiates with the NFLPA.¹⁴⁵ While a settlement that exceeds one billion dollars will not bankrupt the league, the bad press and the shadow that brain injuries cast over the game remain a thorn in the NFL’s side.

138. See *NFL Concussion Claims Hit \$500 Million in Less Than 2 Years*, USA TODAY (July 30, 2018, 2:32 PM), <https://www.usatoday.com/story/sports/nfl/2018/07/30/nfl-concussion-claims-hit-500-million-in-less-than-2-years/37205917/> [<https://perma.cc/56DK-BPPC>].

139. See *id.*

140. See *Concussion Protocol & Return-to-Participation Protocol: Overview*, NFL PLAYER HEALTH & SAFETY, <https://www.nfl.com/playerhealthandsafety/health-and-wellness/player-care/concussion-protocol-return-to-participation-protocol> [<https://perma.cc/B9HN-YALU>] (last visited Oct. 29, 2021).

141. See *Who We Are*, FOOTBALL PLAYERS HEALTH STUDY AT HARVARD UNIV., <https://footballplayershealth.harvard.edu/about/> [<https://perma.cc/EN9A-WAJQ>] (last visited Oct. 29, 2021). The Concussion Protocol is reviewed each year. *Id.*

142. See Deubert et al., *supra* note 132, at 188.

143. See *id.* at 187–89.

144. See Tracey B. Carter, *From Youth Sports to Collegiate Athletics to Professional Leagues: Is There Really “Informed Consent” by Athletes Regarding Sports-Related Concussions?*, 84 UMKC L. REV. 331, 334–35 (2015).

145. See Ashleigh Weinbrecht, Note, *The Connection Between Concussions and Chronic Traumatic Encephalopathy (CTE) in Professional Athletics: A Necessary Change in the Sports Culture in Light of Legal Barriers*, 14 J. L. SOC’Y 309, 327 (2013). The 2020 collective bargaining agreement contains a robust player disability and neurocognitive benefit plan. See *Collective Bargaining Agreement*, NAT’L FOOTBALL LEAGUE PLAYERS’ ASS’N (2020), https://nflpaweb.blob.core.windows.net/media/Default/NFLPA/CBA2020/NFL-NFLPA_CBA_March_5_2020.pdf [<https://perma.cc/33VT-6F6E>]. It also includes a neurocognitive benefit release and covenant not to sue, which covenants players and their families to not pursue claims for head injuries in consideration for benefits provided under the plan. *Id.*

The average NFL career lasts 3.2 years.¹⁴⁶ Similarly, content moderators do not remain in their positions for more than one or two years.¹⁴⁷ Much like the NFL, the technology industry's approach to the issues facing these high-injury-risk, low-tenure workers has been inconsistent application of internal policies and settlement.¹⁴⁸ Unlike NFL players, moderators do not have a strong union protecting their interests.¹⁴⁹ Technology companies do not benefit from the ability to negotiate solutions and forestall potential legal remedies through collective bargaining.

The NFL cannot exist without players. Large platforms cannot exist without moderators, and the issues facing them are not going away anytime soon.¹⁵⁰ Settlements may provide a temporary salve, but they do not heal future wounds. The proposed \$52 million dollar settlement in *Scola* covers moderators working in Arizona, California, Florida, and Texas but will not cover future workers hired by Facebook or its contractors in other states.¹⁵¹ The proposed settlement is notably silent as to foreign outsourced vendors.¹⁵² The settlement's provisions for health care funds do not apply to foreign workers or to future workers.¹⁵³ Without robust worker protections like those available under the NFL's collective bargaining agreement, the same issues will arise with each new crop of content moderators.

2. Pitfalls to Inaction

Both foreign and domestic content moderators fill a role that is akin to factory workers but is masked by the trappings of the information sector.¹⁵⁴ Garment industry practices tell a cautionary tale of the human tragedy and negative political and business consequences that can result from worker neglect and help to illustrate why reforms are needed to better protect workers in the content moderation industry.

In 2013, the collapse of the Rana Plaza building in Dhaka, Bangladesh, killed at least 1132 workers and injured more than 2500.¹⁵⁵ The disaster was

146. See Dashiell Bennett, *The NFL's Official Spin on Average Career Length Is a Joke*, BUS. INSIDER (Apr. 18, 2011, 5:18 PM), <https://www.businessinsider.com/nfls-spin-average-career-length-2011-4> [<https://perma.cc/3X83-JEAJ>] (discussing the NFL's attempt to lengthen the average statistic by removing certain players from their calculations).

147. See ROBERTS, *supra* note 1, at 124–25.

148. See *infra* Part II.B.

149. See *infra* Part II.D.

150. For a discussion of the industry's struggle to solve content moderation issues with artificial intelligence, see *infra* Part II.B.

151. See *supra* note 117 and accompanying text.

152. See *supra* note 117 and accompanying text.

153. See *supra* note 117 and accompanying text. In contrast, the NFLPA agreement, negotiated last year, runs from 2020 to 2030 and will apply to all players who enter the league in the ensuing decade. See NAT'L FOOTBALL LEAGUE PLAYERS' ASS'N, *supra* note 145.

154. See ROBERTS, *supra* note 1, at 59.

155. See *The Rana Plaza Accident and Its Aftermath*, INT'L LABOUR ORG., https://www.ilo.org/global/topics/geip/WCMS_614394/lang--en/index.htm [<https://perma.cc/7Q8X-T9H7>] (last visited Oct. 29, 2021).

among the worst industrial accidents on record, and it united global and local stakeholders to create an unprecedented coordinated framework to compensate the injured workers and dependents of the deceased.¹⁵⁶ Following global uproar, more than 220 apparel companies¹⁵⁷ signed the Accord on Fire and Building Safety in Bangladesh (“Accord”).¹⁵⁸ The Accord was an independent, legally binding agreement that committed signatories to cease business with suppliers who did not submit to inspections, guarantee the rights of workers to refuse unsafe conditions, or perform repairs and renovations, among other stipulations.¹⁵⁹

As content moderation receives increased media attention, as lawsuits accumulate, and as journalists turn a critical eye to working conditions, large technology companies are in the midst of a workplace disaster of their own. Calls for breaking up big technology¹⁶⁰ and for Section 230 reform¹⁶¹ are increasing as the world grapples with technology’s effect on politics and society. Content moderators’ working conditions likely will not culminate in one headline-grabbing tragedy like a building collapse. While an isolated worker suicide or violent side effects of a PTSD episode would be tragic and may gain media attention, a more diffuse crisis is already underway. The hidden nature of content moderation is similar to the supply chain layers that obscure garment workers. The fact that mental health challenges are often invisible also contributes to the slow creep of this silent crisis. While the public may not be watching in horror quite yet, technology companies are fully aware of the mental health risks inherent in content moderation.¹⁶² Maintaining the status quo while a large workforce develops PTSD would be a public relations disaster and would do nothing to stave off eager regulators.

There is little about the big technology companies’ structure or the economics of the labor of content moderation that encourages change.

156. *See id.*

157. *See About, ACCORD ON FIRE AND BLDG. SAFETY IN BANGL.*, <https://bangladeshaccord.org/about> [<https://perma.cc/C9GT-Q6T9>] (last visited Oct. 29, 2021).

158. *See* Jaakko Salminen, *The Accord on Fire and Building Safety in Bangladesh: A New Paradigm for Limiting Buyers’ Liability in Global Supply Chains?*, 66 AM. J. COMPAR. L. 411, 415–17 (2018).

159. *See id.* at 416–17. Many companies that did not want to commit to the binding structure of the Accord joined the Alliance for Bangladesh Worker Safety. *Id.* at 419. In June 2020, the functions of the local office of the Accord transitioned to the RMG Sustainability Council. *See Transition to the RMG Sustainability Council, ACCORD ON FIRE AND BLDG. SAFETY IN BANGL.*, <https://bangladeshaccord.org/updates/2020/06/01/transition-to-the-rmg-sustainability-council-rsc> [<https://perma.cc/5R35-6G5H>] (last visited Oct. 29, 2021).

160. *See* Cecilia Kang et al., *U.S. Accuses Google of Illegally Protecting Monopoly*, N.Y. TIMES (Oct. 20, 2020), <https://www.nytimes.com/2020/10/20/technology/google-antitrust.html> [<https://perma.cc/5SLV-EAYZ>].

161. *See* Daisuke Wakabayashi, *Legal Shield for Social Media Is Targeted by Lawmakers*, N.Y. TIMES (Dec. 15, 2020), <https://www.nytimes.com/2020/05/28/business/section-230-internet-speech.html> [<https://perma.cc/W83D-E46M>] (noting that Democrats have threatened Section 230 reform because the law shields sites from having to tackle problematic content and harassment).

162. *See supra* note 84 and accompanying text; *infra* Part B.

Certain companies deliberately choose to use contractors so that they can argue that workers were only temporary short-term employees when harm is alleged in the future.¹⁶³ While Facebook's \$52 million settlement in *Scola* is nothing for a company making over \$70 billion in annual revenue, the nature of content moderation is such that new workers will enter the market every few years or so.¹⁶⁴ The NFL Concussion Settlement and the Rana Plaza disaster serve as cautionary tales to the technology industry about what happens when you turn a blind eye to worker safety. In addition to the strategic and business risks inherent in maintaining the status quo, increased media coverage¹⁶⁵ and increased public awareness of the vital role content moderators play in society make the status quo untenable.

B. Internal Employment Policies and Corporate Social Responsibility

In February of 2018, academics, corporate lawyers, and content moderators gathered at the Santa Clara University School of Law for the Content Moderation & Removal at Scale conference.¹⁶⁶ In her remarks, Monika Bickert, Facebook's Head of Global Policy Management, stated that companies are a long way from using artificial intelligence to solve the problems inherent in content moderation work.¹⁶⁷ An overarching theme of the conference—for platforms large and small—was that current algorithms are not equipped to make increasingly complex moderation decisions because they cannot determine context.¹⁶⁸ The conference panelists stressed that the most that artificial intelligence can do in the foreseeable future is provide human moderators with more efficient tools.¹⁶⁹ With human

163. See ROBERTS, *supra* note 1, at 126.

164. See *supra* note 92 and accompanying text.

165. See IM SCHATTEN DER NETZWELT, *supra* note 58. *The Cleaners* was an official selection at the 2018 Sundance Film Festival.

166. See *Content Moderation & Removal at Scale*, SANTA CLARA UNIV. SCH. OF L., <https://law.scu.edu/event/content-moderation-removal-at-scale/> [https://perma.cc/43J8-NM7L] (last visited Oct. 29, 2021). The recordings of the Content Moderation & Removal at Scale conference are available at the links cited throughout this section. These recordings are an unparalleled source of candid discussion regarding moderation from lawyers, policymakers, and academics in a field that is often shrouded in secrecy.

167. See Monika Bickert, Head of Glob. Pol'y Mgmt., Facebook, Santa Clara University School of Law Conference: Content Moderation & Removal at Scale (Feb. 2, 2018), <https://santaclarauniversity.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=6e2bf22d-52cd-4e3f-9324-a8810187bad7> [https://perma.cc/7VJB-5SZ7]; see also ROBERTS, *supra* note 1, at 207.

168. See generally Panel on Humans vs. Machines, Santa Clara University School of Law Conference: Content Moderation & Removal at Scale (Feb. 2, 2018), <https://santaclarauniversity.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=e3f266d6-2b5b-4b38-a8e3-a88101179428> [https://perma.cc/RQW7-FG44]. See also Olivier Sylvain, *Recovering Tech's Humanity*, 119 COLUM. L. REV. F. 252, 260–61 (2019).

169. See, e.g., Charlotte Willner, Exec. Dir., Trust & Safety Pro. Ass'n, Panel on Employee/Contractor Hiring, Training, and Mental Well-being, Santa Clara University School of Law Conference: Content Moderation & Removal at Scale (Feb. 2, 2018), <https://santaclarauniversity.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=2a80a263-a73a-48e4-9e76-a88101177f22> [https://perma.cc/TGQ6-6BSV].

moderators here to stay, technology companies must address the plight of content moderators. One way technology companies can improve working conditions is by changing internal and industry-wide policies and measures.

The corporate social responsibility movement began in the 1970s with calls for government intervention to make large corporations and their managers accountable for societal issues.¹⁷⁰ In the 1990s, the movement shifted away from reliance on government regulation toward goals set by corporations and their managers.¹⁷¹ This approach to corporate governance considered not only shareholders but also employees, consumers, local communities, and other groups as stakeholders.¹⁷² As the global community became increasingly aware of environmental and worker health and safety issues, the movement gained momentum.¹⁷³

Critics of the corporate social responsibility movement argue that goals and monitoring schemes replace more rigorous governmental or union oversight of working conditions.¹⁷⁴ From this standpoint, internal schemes are designed not to protect workers or improve their conditions but to limit legal liability and protect against bad press.¹⁷⁵ Critics doubt whether auditors can be trusted to make honest assessments and transparently report their conclusions.¹⁷⁶ In the manufacturing context, suppliers may struggle to implement codes from multiple companies that prioritize different goals.¹⁷⁷

On one level, corporate social responsibility is the natural response to the reality of globalization and the inability of developing states to effectively enforce labor laws and regulations.¹⁷⁸ Yet, Nike's robust compliance program serves as an example of corporate social responsibility's somewhat underwhelming success. Research suggests that in spite of a dedicated compliance staff and program, monitoring of over 800 suppliers resulted in the same or worsened workplace conditions over time for the majority of suppliers.¹⁷⁹ The Nike study indicates that corporate social responsibility cannot function effectively on its own.¹⁸⁰ It must be paired with meaningful monitoring, worker participation, external pressure from states or unions, or preferably all of the above.¹⁸¹

170. See Douglas M. Branson, *Corporate Social Responsibility Redux*, 76 TULANE L. REV. 1207, 1211 (2002).

171. See *id.* at 1217, 1225.

172. See *id.* at 1215.

173. See *id.* at 1225.

174. See Richard M. Locke et al., *Does Monitoring Improve Labor Standards?: Lessons from Nike*, 61 INDUS. & LAB. RELS. REV. 3, 5 (2007).

175. See *id.*; see also James J. Brudney, *Envisioning Enforcement of Freedom of Association Standards in Corporate Codes: A Journey for Sinbad or Sisyphus?*, 33 COMPAR. LAB. L. & POL'Y J. 555, 556 (2012).

176. See Locke et al., *supra* note 174, at 5.

177. See *id.* at 6.

178. See *id.* at 5.

179. See *id.* at 9, 17, 19–20.

180. See *id.* at 22.

181. See generally Brudney, *supra* note 175 (emphasizing the need for freedom of association protections and encouraging greater rule-of-law enforcement of corporate codes).

Whatever its shortcomings may be, the technology sector does not shy away from corporate social responsibility. Environmental sustainability campaigns are frequently extolled by the industry.¹⁸² A sense of corporate social responsibility is partly, although not predominantly, what drives companies to moderate content in the first place.¹⁸³ Many large technology companies have already applied the corporate social responsibility model to address content moderators' working conditions. The question is whether these measures are enough.

1. Company-Wide Measures

An overarching theme in interviews with moderators is that they take their jobs very seriously despite their working conditions.¹⁸⁴ Many express that their jobs would be markedly improved if there were paths to full-time employment with the client companies.¹⁸⁵ The concessions in the settlement agreement in *Scola* highlight what many workers who moderate content are seeking.¹⁸⁶ Policies like making licensed counselors available and creating clear protocols regarding breaks from overwhelming content would be a low-cost, bare-minimum starting point for the industry.¹⁸⁷

Although smaller companies do not face the problems of scale that confront Facebook, Google, or Twitter, their approach could serve as a useful guide for internal approaches that protect workers. Reddit primarily relies on user moderators to remove content, but the company hired the psychiatrist-founded Workplace Wellness Project to help staff cope with viewing and removing illegal content from the platform.¹⁸⁸ At Pinterest, humans review all content that is not spam.¹⁸⁹ The company offers content moderators money to leave after a certain period to prevent burnout in

See also Lance Compa, *Corporate Social Responsibility and Workers' Rights*, 30 COMP. LAB. L. & POL'Y J. 1, 6 (2008) (positing that corporate social responsibility cannot be successful without strong laws enforced by government authorities and democratic trade unions). For a discussion of worker-driven codes, see *infra* Part II.D.2.

182. *See* Amazon News, *Amazon's First Custom 100% Electric Delivery Vehicle*, YOUTUBE (Oct. 8, 2020), <https://www.youtube.com/watch?v=rJJNTVy62PU> [<https://perma.cc/5E2K-LB7T>]; Facebook, *Sustainability*, <https://sustainability.fb.com/> [<https://perma.cc/38KX-8F73>] (last visited Oct. 29, 2021). Many large technology companies are on RepTrak's "Global RepTrak 100" list, which is "the definitive ranking of corporate reputation for the world's leading companies." *See 2021 Global RepTrak 100*, REPTRAK, <https://www.reptrak.com/global-reptrak-100/> [<https://perma.cc/25W9-NV2B>] (last visited Oct. 29, 2021).

183. *See* Klonick, *supra* note 5, at 1616.

184. *See* Newton, *supra* note 11; IM SCHATTEN DER NETZWELT, *supra* note 58 (depicting a Filipino content moderator describing the work, stating, "Our main goal is to make the platform as healthy as possible. We are like policemen. Someone needs to guard it.").

185. *See* ROBERTS, *supra* note 1, at 83; Newton, *supra* note 11.

186. *See supra* note 118 and accompanying text.

187. *See id.*

188. *See* Solon, *supra* note 14; *Our Work*, WORKPLACE WELLNESS PROJECT, <https://theworkplacewellnessproject.com/our-work> [<https://perma.cc/3ZR5-CAAG>] (last visited Oct. 29, 2021).

189. *See* Charlotte Willner, *supra* note 169.

moderators' pursuit of a foot-in-the-door in the technology industry.¹⁹⁰ Pinterest also offers moderators monthly massage benefits and counselor visits every six weeks for group and solo sessions.¹⁹¹ Smaller media companies, like Automattic (the parent company of WordPress) and Medium, recruit content moderators internally from customer support team members who are already passionate about the product and its content.¹⁹² At the Santa Clara conference, Alex Feerst, then-general counsel and head of trust and safety at Medium, stressed that a content moderator must have "the mind of a philosopher . . . the gut of a police detective . . . and the heart of a kindergarten teacher."¹⁹³ The panel on worker health and safety at the conference acknowledged that there are questions about whether content moderation is something that can safely be done as a "career."¹⁹⁴ Panelists stressed the importance of camaraderie in the workplace, along with building strong teams and an awareness among other teams about content moderation's critical role.¹⁹⁵ Panelists acknowledged that the isolated data center model is not sustainable.¹⁹⁶ Notably absent from the panels on worker health and moderation outsourcing were large companies like Facebook and Google.¹⁹⁷

Professor Roberts noted that while the Silicon Valley contract workers she interviewed were working for one of the most successful internet companies of all time, which was notorious for its lush working conditions and endless perks, the content moderators were denied many of these benefits, including health care.¹⁹⁸ These workers were not hired because they "had the mind of a philosopher"—they were impoverished recent college graduates trying to survive in the highly competitive San Francisco job market.¹⁹⁹ One moderator Roberts interviewed expressed frustration that the parent company continued to expand in other areas, like copyright and security, but did not offer a path to full-time employment to temporary contract moderators, even

190. *See id.*

191. *See id.*

192. *See generally* Panel on Employee/Contractor Hiring, Training, and Mental Well-being, Santa Clara University School of Law Conference: Content Moderation & Removal at Scale (Feb. 2, 2018) [hereinafter Panel on Employee/Contractor Hiring, Training, and Mental Well-being], <https://santaclarauniversity.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=2a80a263-a73a-48e4-9e76-a88101177f22> [<https://perma.cc/TGQ6-6BSV>].

193. Alex Feerst, Gen. Couns. at Neuralink, Panel on Employee/Contractor Hiring, Training, and Mental Well-being, Santa Clara University School of Law Conference: Content Moderation & Removal at Scale (Feb. 2, 2018), <https://santaclarauniversity.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=2a80a263-a73a-48e4-9e76-a88101177f22> [<https://perma.cc/TGQ6-6BSV>].

194. *See generally* Panel on Employee/Contractor Hiring, Training, and Mental Well-being, *supra* note 192.

195. *See id.*

196. *See id.*

197. *See id.*

198. *See* ROBERTS, *supra* note 1, at 126.

199. *See id.* at 81.

though moderators already knew the company's systems and policies.²⁰⁰ A content moderator working in a company that has fostered a "team" atmosphere among its moderators, may feel more comfortable seeking counseling. For those in a temporary contract worker position, admitting a need for help may be akin to admitting to not having the skills for the job.²⁰¹

When Facebook was named the number one place to work by the online job search platform Glassdoor in 2018, Facebook Vice President of People Lori Goler expressed that one of her favorite posters in the office reads, "This is your company now."²⁰² Goler gushed that the poster made a statement to workers that they were all in it together.²⁰³ Large technology firms can follow in the steps of their smaller counterparts by extending the "all in this together" mentality to their contract and outsourced moderators whether they are working in the United States, India, the Philippines, or elsewhere.

2. Industry-Wide Measures

Panelists at the Santa Clara conference acknowledged that industry-wide best practices would be helpful but that philosophical differences among the platforms and their products hamper collaboration.²⁰⁴ Additionally, privacy laws and antitrust laws prevent the industry from creating an Interpol-like information sharing network to collectively screen content.²⁰⁵ Despite these challenges, there has been an effort to create universal guidelines, at least for the screening of images of child sexual exploitation.

The Technology Coalition, which counts Amazon, Apple, Snap, Google, Microsoft, and Facebook among its members,²⁰⁶ has promulgated guidelines ("Technology Coalition Guidelines") to support employees who come in contact with images of child sexual exploitation at work.²⁰⁷ The Technology Coalition Guidelines encourage companies to obtain informed consent from employees by describing the content they may encounter and to outline

200. *See id.* at 83.

201. *See id.* at 113; Solon, *supra* note 14.

202. *See* Glassdoor, *Facebook Employees Share Why It Is #1 Best Place to Work in 2018*, YOUTUBE (Dec. 5, 2017), <https://www.youtube.com/watch?v=YTtIYX2XaI> [<https://perma.cc/TLE2-MUBB>].

203. *See id.*

204. *See generally* Panel on Employee/Contractor Hiring, Training, and Mental Well-being, *supra* note 192 (comparing the different approaches to moderation taken by companies like Reddit or Wikipedia, where users drive moderation to companies like Facebook that actively moderate).

205. *See id.*

206. *See* TECH. COAL., <https://www.technologycoalition.org/> [<https://perma.cc/Z2TS-VDSN>] (last visited Oct. 29, 2021).

207. *See generally* *Employee Resilience Guidebook for Handling Child Sexual Abuse Images*, TECH. COAL. (2015), <https://static1.squarespace.com/static/5539d022e4b0a048151fd94b/t/57a820f02e69cffb1a382a7b/1470636275925/TechnologyCoalitionEmployeeResilienceGuidebookV2January2015.pdf> [<https://perma.cc/XWP9-V6QN>]. Industry-wide guidelines would solve the issues garment and other manufacturing suppliers face in applying guidelines from different corporations. *See supra* note 177 and accompanying text.

warning signs they should look for in their own responses to viewing the content at the interview stage.²⁰⁸ When employees start work, they should be exposed to content in a controlled manner with seasoned team members or a counseling service provider.²⁰⁹ Post-exposure counseling sessions should be set up for the new employee.²¹⁰ The guidelines acknowledge that employee resilience programs will not be one-size-fits-all but state the critical need for administration by a nonemployee professional with specialized training in trauma intervention.²¹¹ They provide a long list of potential elements of personal safety plans companies should encourage workers to formulate, including going for a fifteen-minute walk when having a bad reaction, moving on to a different work activity for a period of time, calling a counselor, or taking time off.²¹² While the Technology Coalition Guidelines are specific to images of child sexual exploitation, they acknowledge that content that depicts violence toward people or animals can impact employees.²¹³ They do not contemplate the impact of other images or speech, such as hate speech, propaganda, or conspiracy theory content on moderators.²¹⁴ The Technology Coalition Guidelines are not binding but serve as a reference for member companies.²¹⁵ Despite the existence of these guidelines, moderators allege in court filings that certain companies' contractors do not adhere to them.²¹⁶

The Technology Coalition Guidelines borrow from practices of the National Center for Missing & Exploited Children, the International Centre for Missing & Exploited Children, the Internet Watch Foundation, and the Child Exploitation and Online Protection Centre, which is a U.K. law enforcement agency.²¹⁷ The Internet Watch Foundation is a U.K. nonprofit that works internationally and in partnership with technology companies to remove online images and videos of child abuse.²¹⁸ Potential Internet Watch Foundation analysts are assessed by psychologists and subjected to an enhanced background check.²¹⁹ They undergo six months of training to understand criminal law, learn about the dark web, and build resilience to viewing traumatic content.²²⁰ At the National Center for Missing and Exploited Children, the wellness program extends after analysts leave the

208. See TECH. COAL., *supra* note 207, at 8.

209. See *id.* at 11.

210. See *id.*

211. See *id.* at 12.

212. See *id.* at 14–15.

213. See *id.* at 14.

214. See generally *id.*

215. See generally *id.*

216. See Plaintiffs' Amended Class Action Complaint and Demand for Jury Trial, *supra* note 53, ¶¶ 49–59.

217. See TECH. COAL., *supra* note 207, at 11.

218. See generally *What We Do*, INTERNET WATCH FOUND., <https://www.iwf.org.uk/what-we-do> [<https://perma.cc/L3KA-TRPC>] (last visited Oct. 29, 2021).

219. See Solon, *supra* note 14.

220. See *id.*

organization and offers support for analysts' spouses and significant others.²²¹ Large technology companies have started to look to these organizations as a model but could expand on the suggestions of the Technology Coalition by making the guidelines mandatory, by creating guidelines that apply to all types of content, or by creating industry-wide minimum worker wellness program standards.

The Technology Coalition Guidelines draw from studies evaluating the effect of exposure to disturbing media on law enforcement officers.²²² There are no similar studies to track content moderator wellness over their period of employment or beyond.²²³ Panelists at the Santa Clara conference acknowledged that while there are parallels between counseling given to intensive care unit staff, emergency medical technicians, or members of the military, this often does not encompass what reviewers are experiencing when they sit at a desk and experience trauma all day long, rather than in isolated incidents.²²⁴ Increased research into the effects content has on workers will be essential to creating industry-wide programs.

3. Recommendations from the Experts

“[C]ontent moderation seems so important to running Facebook that it ought to be regarded as falling within the company’s core activities, not as an ancillary chore to be handled by contractors.”²²⁵ This observation by Paul Barrett, Deputy Director of the New York University Stern Center for Business and Human Rights, in his 2020 report on content moderation, is typical of the sentiment of most who study content moderation work closely.²²⁶ Barrett’s recommendations, and those of others who have studied content moderation in both academia and through journalistic investigations, provide suggestions that span the realm of both internal and collective change.

Barrett’s report offers three principal recommendations. First, Barrett argues that technology companies should end outsourcing in content moderation completely.²²⁷ He does not suggest that all moderators should

221. *See id.*

222. *See* TECH. COAL., *supra* note 207, at 7–10.

223. *See* ROBERTS, *supra* note 1, at 209; Sarah Roberts, *Commercial Content Moderation and Worker Wellness: Challenges & Opportunities*, TECHDIRT (Feb. 8, 2018, 1:56 PM), <https://www.techdirt.com/articles/20180206/10435939168/commercial-content-moderation-worker-wellness-challenges-opportunities.shtml> [<https://perma.cc/6KT2-P3YJ>].

224. *See generally* Panel on Employee/Contractor Hiring, Training, and Mental Well-being, *supra* note 192.

225. *See* BARRETT, *supra* note 20, at 18.

226. *See* ROBERTS, *supra* note 1, at 82; Newton, *supra* note 11.

227. *See* BARRETT, *supra* note 20, at 24. Barrett’s suggestion underscores a larger national debate regarding independent contractor labor. *See, e.g.*, *Dynamex Operations W., Inc. v. Sup. Ct. of L.A. Cnty.*, 416 P.3d 1 (Cal. 2018); Assemb. Bill 5, 2019–2020 Reg. Sess., ch. 296 (Cal. 2011) (amending section 3351 and adding to section 2750.3 of the Labor Code, amending sections 606.5 and 621 of the Unemployment Insurance Code, and codifying the *Dynamex* decision and seeking to classify gig workers as employees rather than independent

be employed in the United States, but he suggests they should be “full-fledged employees,” no matter where they are located.²²⁸ Moderators everywhere should look directly to Silicon Valley for supervision, compensation, and overall office well-being.²²⁹ Barrett also suggests that all moderators should have access to top-quality on-site medical care.²³⁰ Health care plans should extend for a period of years after moderators leave the job or until they obtain coverage via a different employer.²³¹ Lastly, Barrett recommends that companies pool resources to sponsor high-quality academic research into the risks inherent in the work of content moderation.²³²

Casey Newton, whose 2019 investigation of Facebook content moderation sites run by Cognizant in Tampa and Phoenix exposed shocking working conditions, also suggests that paths to full-time employment would benefit moderators tremendously.²³³ Newton recommends that the risk of developing PTSD should be disclosed in the job description.²³⁴ He suggests that companies conduct research about which roles and content pose the highest risks for workers and that companies determine a lifetime cap for exposure to harmful content.²³⁵ He also recommends that mental health support should be available even after workers leave their positions.²³⁶

In her book *Behind the Screen*, Professor Sarah Roberts predicts that the workforce of content moderators will only increase as platforms continue to turn to human beings to meet their moderation needs.²³⁷ Roberts expresses doubt that the social media industry will decide to self-regulate to solve the plight of content moderators.²³⁸ *Behind the Screen* concludes with a series of potential avenues for relief for content moderators but does not advocate for any particular measures.²³⁹ However, Roberts suggests that industry-wide best practices, like the Technology Coalition Guidelines or setting time limits on the number of hours worked, could prove to be

contractors); *California Proposition 22, App-Based Drivers as Contractors and Labor Policies Initiative* (2020), BALLOTPEDIA
[https://ballotpedia.org/California_Proposition_22,_App-Based_Drivers_as_Contractors_and_Labor_Policies_Initiative_\(2020\)](https://ballotpedia.org/California_Proposition_22,_App-Based_Drivers_as_Contractors_and_Labor_Policies_Initiative_(2020)) [https://perma.cc/ES5U-XWP9] (last visited Oct. 29, 2021) (noting the success of a referendum rejecting Assembly Bill No. 5 and classifying app-based drivers as contractors).

228. See BARRETT, *supra* note 20, at 24.

229. See *id.* at 2.

230. See *id.* at 25.

231. See *id.*

232. See *id.*

233. See Casey Newton, *What Tech Companies Should Do About Their Content Moderators' PTSD*, VERGE (Jan. 28, 2020, 6:00 AM), <https://www.theverge.com/interface/2020/1/28/21082642/content-moderator-ptsd-facebook-youtube-accenture-solutions> [https://perma.cc/P8QD-HMBF].

234. See *id.*

235. See *id.*

236. See *id.*

237. See ROBERTS, *supra* note 1, at 207.

238. See *id.* at 211.

239. See *id.* at 219–20.

effective.²⁴⁰ Roberts quotes a professional Amazon Mechanical Turk moderator who, when asked what would improve her quality of life as a moderator, simply responded, “Pay us.”²⁴¹ This suggests that the solution could lie in properly compensating moderators in line with the responsibility they take on.²⁴² Roberts stresses her hope that, by exposing the work of content moderators, both technology companies and the general public will understand the true costs of our use of digital platforms.²⁴³

Corporate social responsibility is likely not the magic bullet that will cure the ills in the content moderation industry. However, many companies large and small have taken steps toward improving worker conditions and creating industry-wide codes. Increased user awareness, journalistic exposure, and potential liability could prompt technology companies to begin a more robust, industry-wide corporate social responsibility endeavor that improves working conditions for content moderators.

C. Regulatory and Legislative Solutions

Greater understanding of the impact of technology on our lives has led to an increase in calls for government regulation of large platforms.²⁴⁴ The Trump administration endorsed increased regulation of large technology companies and reform to Section 230 of the Communications Decency Act.²⁴⁵ In October 2020, the U.S. Department of Justice brought suit against Google alleging violation of antitrust laws.²⁴⁶ The Federal Trade Commission and more than forty states sued Facebook, seeking to break up the company.²⁴⁷ While these efforts speak more to squabbles over content, what the platforms do with it, and how they profit from it, rather than the work of content moderators in particular, increased government regulation of the technology industry could impact their work significantly. Recent changes in international law highlight that potential impact. On the other hand, content moderators’ salvation may not lie in new legislation or regulations but in the application of existing schemes to their situation.

240. *See id.* at 210–11.

241. *See id.* at 216–17.

242. *See id.*

243. *See id.* at 219.

244. *See id.* at 212–13; Human Rights Council, Rep. of the Indep. Int’l Fact-Finding Mission on Myanmar, U.N. Doc. A/HRC/39/64, ¶ 74 (Sept. 12, 2018) (“The extent to which Facebook posts and messages have led to real-world discrimination and violence must be independently and thoroughly examined.”).

245. *See* Exec. Order No. 13,925, 85 Fed. Reg. 37,635 (June 2, 2020); Press Release, Dep’t of Just., Off. of Pub. Affs., The Justice Department Unveils Proposed Section 230 Legislation (Sept. 23, 2020), <https://www.justice.gov/opa/pr/justice-department-unveils-proposed-section-230-legislation> [<https://perma.cc/ET2G-TXZF>].

246. Complaint, United States v. Google LLC, No. 1:20-cv-03010 (D.D.C. Oct. 20, 2020).

247. *See* Cecilia Kang & Mike Isaac, *U.S. and States Say Facebook Illegally Crushed Competition*, N.Y. TIMES (July 28, 2021), <https://www.nytimes.com/2020/12/09/technology/facebook-antitrust-monopoly.html> [<https://perma.cc/TBX8-ALG3>].

1. Existing Domestic Legislation and Regulation

Currently, Section 230 immunizes online platforms from liability for content to which they do not materially contribute.²⁴⁸ Recent calls for reform have come from both sides of the political divide.²⁴⁹ Conservatives argue that the text of Section 230 requires political neutrality in moderation.²⁵⁰ Others argue that the doctrine allows platform designs that facilitate discrimination against users from historically marginalized groups and should be amended to include protections for those groups.²⁵¹ Were Section 230 amended in favor of either group, there would likely be an even greater need for human moderators. There would be either more borderline content to analyze or more content that potentially violates platform guidelines. Were Section 230 immunity to be eliminated, as both former President Donald Trump²⁵² and President Joe Biden have advocated,²⁵³ the need for content moderators would explode as the internet platforms could potentially be held liable for user-generated content. Professor Olivier Sylvain has noted that courts are increasingly taking an approach that contemplates the platforms' commercial mission rather than their role in promoting user speech, which could be a clue that we are moving further away from the notions of free and open internet as we grapple with the realities of social media's effects on society.²⁵⁴ While the future of Section 230 presents important questions about the fate of content moderators, protections from the workplace and health issues they are now dealing with are not contemplated in calls for reform.

The scope of existing worker protections for content moderators in the United States could be broadened, either through interpretation by the courts or amendments to existing legislation and regulations. For instance, the Occupational Safety and Health Act of 1970²⁵⁵ ("OSH Act") requires that each employer furnish a workplace free from recognized hazards that are causing or are likely to cause death or serious physical harm to their employees.²⁵⁶ The clause requires that employers discover and exclude all feasibly preventable forms of hazardous conduct from the workplace.²⁵⁷ A

248. See Sylvain, *supra* note 168, at 269.

249. See *id.* at 270, 274.

250. See *id.* at 270.

251. See *id.* at 273–74.

252. See Derek E. Bambauer, *Trump's Section 230 Reform Is Repudiation in Disguise*, BROOKINGS (Oct. 8, 2020), <https://www.brookings.edu/techstream/trumps-section-230-reform-is-repudiation-in-disguise/> [<https://perma.cc/P2Q5-YXU7>].

253. See Makena Kelly, *Joe Biden Wants to Revoke Section 230*, VERGE (Jan. 17, 2020, 10:29 AM), <https://www.theverge.com/2020/1/17/21070403/joe-biden-president-election-section-230-communications-decency-act-revoke> [<https://perma.cc/569V-242J>].

254. See Sylvain, *supra* note 168, at 279–80.

255. Pub. L. No. 91-596, 84 Stat. 1590 (codified as amended in scattered sections of the U.S.C.).

256. 29 U.S.C. § 654 (a)(1).

257. See *Nat'l Realty & Constr. Co. v. Occupational Safety and Health Rev. Comm'n*, 489 F.2d 1257, 1267 (D.C. Cir. 1973).

review board or court must determine whether a precaution is recognized by safety experts as feasible.²⁵⁸ Precautions do not become infeasible merely because they are expensive.²⁵⁹ In the OSH Act's early years, the general duty clause served as a stopgap to prohibit hazardous conditions before specific standards became effective.²⁶⁰ The courts have interpreted the clause to cover peculiar violations not covered by specific standards.²⁶¹ While the text of the general duty clause clearly covers "serious physical harm," which may not literally encompass the mental health effects content moderators experience, the Act's legislative history and subsequent application suggest that it could be extended to the physical manifestations of moderators' stress.²⁶²

The Occupational Safety and Health Administration's (OSHA) Critical Incident Stress Guide acknowledges that there are no standards that apply to the hazards associated with critical incident stress, which occurs after witnessing or experiencing tragedy, death, serious injuries, or threatening situations.²⁶³ The guide outlines strategies to help employers reduce stress and conduct critical incident stress debriefings²⁶⁴ but clearly does not contemplate that a worker's entire job may entail exposure to critical incidents. If OSHA review boards and courts interpret "physical harm" broadly, content moderators could find relief via OSHA complaints. OSHA itself could pass additional regulations that cover critical incident stress and content moderation work. Or, as Congress considers greater regulation in the technology space, it could amend the OSH Act to include mental harm. At the same time, the agency's extremely weak response to COVID-19-related workplace complaints highlights the limitations of relying on OSHA for relief when there is a lack of political will within the agency to take action.²⁶⁵

State workers' compensation schemes could also potentially provide relief to individual workers diagnosed with mental health conditions as a result of

258. *See id.* at 1266 n.37.

259. *See id.*

260. *See* 3 WEST'S FED. ADMIN. PRAC. § 2617 (2021).

261. *See* Sec'y of Lab. v. S. Soya Corp., 5 OSAHRC 309 (No. 515, 1973) (holding that a manager's order for an employee to enter a cotton storage tank where he was subsequently injured violated the general duty clause).

262. *See* Newton, *supra* note 46 (discussing the fatal heart attack a content moderator suffered at Cognizant's Tampa site). The stopgap nature of the clause itself could provide some protection for workers. Legislative history suggests that the goal of the clause was to ensure that employees working under special circumstances without adopted standards would be protected. *See* S. REP. NO. 91-1282, at 5186 (1970) (Conf. Rep.). Additionally, the drafters considered "occupational disease" a major harm targeted by the OSH Act. In 1970, Congress was concerned with cancer, respiratory ailments, and allergies among other diseases. *See id.* at 5178. The specific intent and larger purpose of this provision could be expanded to include the mental health "diseases" content moderators experience.

263. *See Critical Incident Stress Guide*, OCCUPATIONAL SAFETY & HEALTH ADMIN., <https://www.osha.gov/emergency-preparedness/guides/critical-incident-stress> [<https://perma.cc/ZW53-62AH>] (last visited Oct. 29, 2021).

264. *See id.*

265. *See generally* James J. Brudney, *Forsaken Heroes: COVID-19 and Frontline Essential Workers*, 48 FORDHAM URB. L.J. 1 (2020).

their content moderation work. There is variance among state workers' compensation law as to whether mental injuries are compensable.²⁶⁶ Some states generally cover mental-only injuries without a physical stimulus, like PTSD; others cover what are known as "mental-mental" injuries only in limited circumstances involving a sudden stimulus or carve out exceptions for first responders.²⁶⁷ A group of states does not cover mental-only injuries at all.²⁶⁸ Even in states that do compensate so called "mental-mental" claims, courts may raise the bar and require claimants to show that the injury was caused by "unusual stress" that was beyond the day-to-day emotional strain and tension experienced by all employees.²⁶⁹ Content moderators whose very job description encompasses what others would think of as day-to-day emotional strain would be unlikely to meet this standard.

The workers who have filed suit represent a useful sample for analyzing the availability of workers' compensation for content moderators. In California, where Selena Scola worked, an employee may not recover for a psychiatric injury resulting from a regular and routine employment event, unless the employee has worked for the employer for six months.²⁷⁰ Six months, in the work history of a content moderator who will only stay in the position for a year or two, is a long period of time. Moderators who became distressed early on in their tenure would have difficulty obtaining compensation in California. Florida, where Debryna Garrett was moderating content for Facebook, provides no relief for purely mental workers' compensation claims.²⁷¹ In Arizona, mental injury, illness, or condition is not covered by workmen's compensation, unless some unexpected stress related to the employment or a physical injury related to the employment was a substantial contributing cause.²⁷² Greg Soto's Washington state workers' compensation claim was denied.²⁷³ In Washington, claims based on mental conditions or mental disabilities caused by stress are not covered unless they relate to sudden stress resulting from a single traumatic event, with exceptions for first responders.²⁷⁴ Were all states to adopt coverage for "mental-mental" injuries, workers' compensation could provide relief for content moderators. However, the inconsistency among state laws leaves them vulnerable.

266. *See generally* 4 LARSON'S WORKERS' COMPENSATION LAW § 56.04 (2020).

267. *See id.*

268. *See id.*

269. *See* *United Parcel Serv., Inc. v. Lust*, 560 N.W.2d 301, 305 (Wis. Ct. App. 1997) (citing *Sch. Dist. No. 1 v. DILHR*, 215 N.W.2d 301 (Wis. 1974)).

270. *See Hansen v. Workers' Comp. Appeals Bd.*, 18 Cal. App. 4th 1179, 1182 (1993).

271. *See* 4 LARSON'S WORKERS' COMPENSATION LAW § 56.06 (2020).

272. *See* ARIZ. REV. STAT. ANN. § 23-1043.01(B) (2021).

273. *See supra* note 105 and accompanying text.

274. *See* WASH. REV. CODE § 51.08.142 (2021); WASH. ADMIN. CODE § 296-14-300 (2021).

2. Potential New Domestic Legislation

There is potential for Congress to consider content moderation as its own issue meriting particular legislation or regulation, rather than attempting to fit content moderators into existing schemes. Congress has considered the particular issues facing content moderators before. In 2010, the Online Safety and Technology Working Group was convened²⁷⁵ pursuant to the Protecting Children in the 21st Century Act.²⁷⁶ The working group's report recommended that the federal government consider incentives for service providers to create wellness programs for employees tasked with viewing disturbing imagery.²⁷⁷ Congress did not implement the working group's recommendation, but the fact that such provisions were considered indicates potential for legislation with nationwide scope. Legislation and/or regulations that protect content moderators nationally could be a solution that not only benefits workers, but creates consistency among the large technology firms.

An analogous example of Congress stepping in to protect victims of particular trauma is the James Zadroga 9/11 Health and Compensation Act of 2010 ("Zadroga Act").²⁷⁸ The Zadroga Act created the World Trade Center Health Program, which provides medical monitoring and treatment benefits to 9/11 first responders and cleanup workers.²⁷⁹ The Zadroga Act also created the September 11th Victim Compensation Fund of 2001.²⁸⁰ The program provides monitoring and treatment for specific health conditions determined to be 9/11-related, including PTSD.²⁸¹ While content moderators are not victims of a particular national tragedy like the 9/11 first responders, they experience shared trauma and provide a valuable public service. Their work protects all internet users, which raises the question of whether or not it is the government's responsibility to care for them. Unlike 9/11 first responders who are a fixed group, content moderators are an expanding group. Hesitation about the government and taxpayers bearing the burden of protecting an ever-growing group could be overcome as we move closer to

275. See generally *Online Safety and Technology Working Group*, NAT'L TELECOMMS. & INFO. ADMIN., <https://www.ntia.doc.gov/legacy/advisory/onlinesafety/index.html> [https://perma.cc/2727-QPBW] (last visited Oct. 29, 2021).

276. See 15 U.S.C. § 6554.

277. See ONLINE SAFETY AND TECH. WORKING GRP., YOUTH SAFETY ON A LIVING INTERNET: REPORT OF THE ONLINE SAFETY AND TECHNOLOGY WORKING GROUP 8 (2010), https://www.ntia.doc.gov/files/ntia/publications/ostwg_final_report_070610.pdf [https://perma.cc/37A5-HYHU].

278. Pub. L. No. 111-347, 124 Stat. 3623 (2011) (codified as amended in scattered sections of the U.S.C.); see 42 U.S.C. § 300mm.

279. See 42 U.S.C. § 300mm.

280. See *id.* § 300mm-61.

281. See *World Trade Center Health Program*, CTRS. FOR DISEASE CONTROL & PREVENTION, <https://www.cdc.gov/wtc/about.html> [https://perma.cc/WAZ8-4JRV] (last visited Oct. 29, 2021).

the view that the internet and social media platforms are spaces akin to public town squares.²⁸²

On a smaller scale, the U.S. Department of Justice has stepped in to provide training programs for law enforcement officers, forensic analysts, prosecutors, judges, and other professionals who have to view child sexual abuse images.²⁸³ Given the fact that content moderators are now often the first line of defense in this work, the idea that the government could subsidize their care or training is not far-fetched.

3. International Legislation

The need for content moderators continues to increase as platforms expand into more countries and support more languages. As the scale and reach of platforms grow, international law increasingly impacts content moderation and could provide avenues for worker protection.

a. Legislation Targeting User Safety

Growing concern about hate speech, misinformation, and other harmful content has spurred the adoption of online safety legislation in European countries. In her Political Guidelines for the Next European Commission, then-candidate for president of the European Commission, Ursula von der Leyen stressed that digital platforms should not be used to destabilize democracy and that the European Union should develop common standards for disinformation and online hate messages.²⁸⁴ Those goals coalesced into the Digital Services Act package, which would impose legal obligations on digital platforms to address the risks faced by users and protect their rights.²⁸⁵ The Digital Services Act package would subject platforms to mandatory notice-and-takedown orders forcing them to remove illegal content, including racism and xenophobia, or face fines.²⁸⁶ Public consultation on the package ended in September 2020.²⁸⁷

282. See *Packingham v. North Carolina*, 137 S. Ct. 1730, 1737 (2017) (“These websites can provide perhaps the most powerful mechanisms available to a . . . citizen to make his or her voice heard.”); Exec. Order No. 13,925, 85 Fed. Reg. 37, 635 (June 2, 2020) (“[T]hese platforms function in many ways as a 21st century equivalent of the public square.”).

283. See SHIFT WELLNESS, <https://www.shiftwellness.org/> [https://perma.cc/CR66-NMDB] (last visited Oct. 29, 2021).

284. See URSULA VON DER LEYEN, EUR. COMM’N, POLITICAL GUIDELINES FOR THE NEXT EUROPEAN COMMISSION 2019–2024 (2019), https://ec.europa.eu/info/sites/info/files/political-guidelines-next-commission_en_0.pdf [https://perma.cc/GCP8-HKJJ].

285. See *The Digital Services Act Package*, EUR. COMM’N, <https://ec.europa.eu/digital-single-market/en/digital-services-act-package> [https://perma.cc/ZN6P-EJM8] (last visited Oct. 29, 2021).

286. See Mehreen Khan & Madhumita Murgia, *EU Draws Up Sweeping Rules to Curb Illegal Online Content*, FIN. TIMES (July 23, 2019), <https://www.ft.com/content/e9aa1ed4-ad35-11e9-8030-530adfa879c2> [https://perma.cc/7BGZ-A8PA].

287. See *The Digital Services Act Package*, *supra* note 285.

While the European Union works toward formulating common standards applicable to platforms in all EU countries, individual countries have acted on their own to regulate online content. In 2017, the German parliament passed the “NetzDG” Act to Improve Enforcement of the Law in Social Networks.²⁸⁸ NetzDG requires platforms to immediately take notice of content reported to them by users and examine whether that content might violate criminal law.²⁸⁹ Platforms must take down or block access to manifestly unlawful content within twenty-four hours of the complaint.²⁹⁰ Other criminal content must generally be taken down or blocked within seven days of receiving a complaint.²⁹¹ A fine of up to fifty million euros can be imposed if a company fails to comply with its obligations under the Act.²⁹² In 2019, France followed suit with Proposition de Loi visant à lutter contre la haine sur internet.²⁹³ Under the proposition, platforms would have twenty-four hours to analyze content flagged as hate speech, including messages attacking someone on the basis of race, religion, sexual orientation, nationality, gender identity, or disability, and propaganda tied to terrorism or war crimes and harassment.²⁹⁴ If the platform refuses to remove such content, it is subject to fines of over one million euros.²⁹⁵ The proposition passed in 2020, but critical provisions were struck down by the French Constitutional Council.²⁹⁶ The court ruled that the obligations in the law created an incentive for platforms to remove flagged content whether or not it was hate speech, thus infringing on freedom of expression and communication.²⁹⁷

Legislators in the United Kingdom have proposed the Online Harms White Paper, which goes a step further than legislation in Germany and France by establishing a new statutory duty of care to hold platforms responsible for the

288. See *Netzwerkdurchsetzungsgesetz [NetzDG] [Network Enforcement Act]*, Sept. 1, 2017, (Ger.), <http://www.gesetze-im-internet.de/netzdg/> [<https://perma.cc/Q83P-9CZ7>].

289. See *id.*

290. See *id.*

291. See *id.*

292. See *id.* In 2019, Facebook was fined two million euros for failing to accurately report the number of hate speech complaints received. See Jenny Gesley, *Germany: Facebook Found in Violation of “Anti-Fake News” Law*, LIBR. OF CONG. (Aug. 20, 2019), <https://www.loc.gov/item/global-legal-monitor/2019-08-20/germany-facebook-found-in-violation-of-anti-fake-news-law/> [<https://perma.cc/9UMF-ALKN>].

293. See Proposition de loi visant à lutter contre la haine sur internet du 19 juin 2019, [Proposal of Law to Fight Against the Hatred on the Internet] Assemblée Nationale, http://www.assemblee-nationale.fr/dyn/15/textes/115b2062_texte-adopté-commission [<https://perma.cc/VCA7-7EU7>] (last visited Oct. 29, 2021).

294. See Aurelien Breeden, *France Will Debate a Bill to Stop Online Hate Speech. What’s at Stake?*, N.Y. TIMES (July 1, 2019), <https://www.nytimes.com/2019/07/01/world/europe/france-bill-to-stop-online-hate-speech.html> [<https://perma.cc/VY4Y-H7UF>].

295. See *id.*

296. See Aurelien Breeden, *French Court Strikes Down Most of Online Hate Speech Law*, N.Y. TIMES (June 18, 2020), <https://www.nytimes.com/2020/06/18/world/europe/france-internet-hate-speech-regulation.html> [<https://perma.cc/XM9B-R956>].

297. See *id.*

safety of their users.²⁹⁸ Compliance with this duty of care would be overseen by an independent regulator who has the power to issue substantial fines and impose liability on individual members of senior management.²⁹⁹ The Irish Online Safety and Media Regulation Bill, which is also in the drafting stage, seeks to establish a Media Commission which would oversee a new regulatory framework for online safety.³⁰⁰ The Media Commission would have the power to impose financial sanctions on companies that do not comply.³⁰¹

European efforts to stem the tide of hate speech and other harmful content will increase the need for content moderators. As the rejection of the French law by the French Constitutional Council shows, schemes such as these will create even greater demand for human moderation if there is pushback on language that would call for automatic takedowns as infringing on freedom of expression. When laws are tempered to address freedom of expression concerns, the need for human moderation becomes even greater. These legislative efforts are also notable for their complete failure to address the plight of content moderators. They would seem to be the perfect vehicle to impose regulation on the working conditions of content moderators, but they are silent on this issue. The Online Harms White Paper mentions “frontline service workers” but only in relation to the need for these workers to provide better support to users.³⁰²

b. Business Incentives

The creation of “special industrial zones” or “special economic zones,” in East Asia in particular, provides favorable terms to transnational corporations or local contract corporations working to serve them.³⁰³ These zones have different terms of governance than traditional sovereign nations and attract businesses by offering tax exemptions and other favorable economic incentives.³⁰⁴ These terms may also include relaxed labor laws, which leave content moderators with little means of redress.³⁰⁵ While an unlikely solution, countries that create special economic zones could build in worker safety requirements for content moderators as part of the cost of doing business in the sector.

298. See *Online Harms White Paper: Full Government Response to the Consultation*, DEP’T FOR DIGIT., CULTURE, MEDIA & SPORT (Dec. 15, 2020), <https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response> [https://perma.cc/WE6J-T54V].

299. See *id.*

300. See *Online Safety & Media Regulation Bill*, DEP’T OF TOURISM, CULTURE, ARTS, GAELTACHT, SPORT & MEDIA (June 2, 2021), <https://www.gov.ie/en/publication/d8e4c-online-safety-and-media-regulation-bill/> [https://perma.cc/34KP-39CG].

301. See *id.*

302. See DEP’T FOR DIGIT., CULTURE, MEDIA & SPORT, *supra* note 298.

303. See ROBERTS, *supra* note 1, at 61–62.

304. See *id.*

305. See *id.*

The Philippines is taking a small step in this direction. There are four bills pending in the House of Representatives of the Philippines that directly address working conditions for business process outsourcing (BPO) workers, a group that broadly captures content moderators.³⁰⁶ The bills were criticized by the National Economic and Development Authority as largely being comprised of “protections” that are already present in the Philippine Constitution, Labor Code, and Occupational Health and Safety Standards.³⁰⁷ Only one of the bills directly addresses the working conditions of content moderators and calls for companies to provide 24/7 on-staff psychologists for workers who are exposed to obscene and violent content.³⁰⁸

c. Transnational Organizations

The International Labour Organization (ILO) is a specialized agency of the United Nations with 187 member states.³⁰⁹ The ILO works to set labor standards, develop policies, and devise programs that promote decent work.³¹⁰ In 2019, the ILO’s Global Commission on the Future of Work released a report calling for a universal labor guarantee of humane working conditions, including protections against sickness, disease, and injury arising out of employment.³¹¹ The organization addressed content moderators in particular in a 2018 report on digital microlabor platforms, such as Amazon Mechanical Turk and Clickworker.³¹² The report discussed the use of digital labor platforms to distribute content moderation work and the long-term psychological toll that viewing violent content can have on workers and contractors in particular.³¹³ The report suggested that it is not always clear

306. See An Act Providing for the Protection of Employees in the Business Process Outsourcing (BPO) Industry, House Bill No. 6190 (Feb. 5, 2020) (Phil.); An Act Ensuring the Welfare and Protection of Business Process Outsourcing (BPO) Workers and the Recognition of Their Rights as Provided for in the Labor Code of the Philippines, House Bill No. 5754 (Dec. 9, 2019) (Phil.); An Act Providing for the Protection of Workers in the Business Process Outsourcing (BPO) Industry, House Bill No. 4236 (Aug. 28, 2019) (Phil.); An Act Ensuring the Welfare and Protection of Business Process Outsourcing (BPO) Workers in the Philippines, House Bill No. 0274 (July 1, 2019) (Phil.); *Position Paper on House Bills on the Protection and Welfare of BPO Workers*, BPO INDUS. EMPS. NETWORK (Aug. 4, 2020), <https://bienphilippines.files.wordpress.com/2020/08/bien-position-paper-on-house-bills-on-the-bpo-employees-protection-and-welfare-final.pdf> [<https://perma.cc/2FJ9-GU6R>].

307. See Charissa Luci-Atienza, *NEDA Asks Lawmakers to Review Policy Gaps in BPO Industry*, MANILA BULL. (July 8, 2020, 12:44 PM), <https://mb.com.ph/2020/07/08/neda-asks-lawmakers-to-review-policy-gaps-in-bpo-industry/> [<https://perma.cc/J7BE-D2MZ>].

308. An Act Ensuring the Welfare and Protection of Business Process Outsourcing (BPO) Workers and the Recognition of Their Rights as Provided for in the Labor Code of the Philippines, House Bill No. 5754, § 22(c) (Dec. 9, 2019) (Phil.).

309. See *About the ILO*, INT’L LABOUR ORG., <https://www.ilo.org/global/about-the-ilo/lang-en/index.htm> [<https://perma.cc/E2P5-G7W3>] (last visited Oct. 29, 2021).

310. See *id.*

311. See INT’L LABOUR ORG., GLOBAL COMM’N ON THE FUTURE OF WORK, WORK FOR A BRIGHTER FUTURE (2019), https://www.ilo.org/wcmsp5/groups/public/---dgreports/---cabinet/documents/publication/wcms_662410.pdf [<https://perma.cc/FH45-2FPG>].

312. See BERG ET AL., *supra* note 45, at 13.

313. See *id.* at 85–86.

on the platforms that particular tasks will entail viewing explicit or offensive content and also suggested that terms of service documents for microlabor platforms should be worker-friendly and include provisions that are respectful of workers' psychological burdens.³¹⁴ The report also recommended including provisions to give workers performing content moderation work access to counseling paid for by the client or platform.³¹⁵

While the ILO's work draws attention to and promotes greater understanding of global working conditions, the organization lacks enforcement power and relies on moral persuasion, publicity, and shame to urge compliance with its conventions.³¹⁶ Conventions and core labor standards do not apply universally, as many are not ratified by all member states.³¹⁷ The United States, for example, has only ratified fourteen of 189 ILO Conventions.³¹⁸ And it has only ratified two of the ILO's core labor standards—those abolishing forced labor and eliminating the worst forms of child labor.³¹⁹

Despite these drawbacks, regulations that are universally or almost universally accepted provide benchmarks for acceptable human rights standards in the workplace globally.³²⁰ For example, content moderators screening out instances of child sexual exploitation are acting consistent with an ILO convention prohibiting the worst forms of child labor, a convention ratified by all 187 ILO member nations, including the United States.³²¹ More generally, the normative force of ILO Standards and their acceptance by different international bodies looms large when employers attempt to advance their efforts intranationally.³²²

Still, content moderators represent a much smaller group than children who are forced to participate in the worst forms of child labor. Given the lack of universal adoption of other ILO conventions, the difficulties in enforcement, and the increasing global trend toward isolationist policies,³²³

314. *See id.* at 103–04.

315. *See id.* at 104.

316. *See Developments in the Law: Jobs and Borders*, 118 HARV. L. REV. 2171, 2207 (2005).

317. *See id.* at 2206–07.

318. *See The US: A Leading Role in the ILO*, INT'L LABOUR ORG., <https://www.ilo.org/washington/ilo-and-the-united-states/the-usa-leading-role-in-the-ilo/lang--en/index.htm> [<https://perma.cc/92X3-GSMJ>] (last visited Oct. 29, 2021).

319. *See* Janice R. Bellace, *ILO Declaration of Fundamental Principles and Rights at Work*, 17 INT'L J. COMPAR. LAB. L. & INDUS. RELS. 269, 279 n.6 (2001).

320. *See* Brudney, *supra* note 175, at 560.

321. *See* Worst Forms of Child Labor Convention, 1999 (ILO No. 182), art.3(b). The ILO's NORMLEX system includes a list of ratifying countries. *See Ratifications of C182–Worst Forms of Child Labour Convention*, ILO, <https://www.ilo.org/dyn/normlex/en/?p=1000:11001::NO::> [<https://perma.cc/RS4L-NTLL>] (last visited Oct. 29, 2021).

322. *See* Tonia Novitz, *Internationally Recognized Right to Strike: A Past, Present, and Future Basis upon Which to Evaluate Remedies for Unlawful Collective Action?*, 30 INT'L J. COMPAR. LAB. & INDUS. RELS. 357, 361 (2014).

323. *See* Stephen Castle, *U.K. Lawmakers Give Brexit Bill the Green Light*, N.Y. TIMES (Nov. 9, 2020), <https://www.nytimes.com/2020/01/09/world/europe/brexit-boris-johnson-parliament-vote.html> [<https://perma.cc/5J3S-BCP7>]; Katie Rogers & Apoorva Mandavilli,

it is unlikely that ILO conventions or measures promulgated by other transnational organizations will specifically target content moderation work.

D. Unionization, Organization, and Worker Mobilization

The barriers to unionization for content moderators are high. Workers are separated geographically and have different statuses of employment and job titles, and their labor is subject to different laws.³²⁴ An international union of content moderators would not be an easy feat to achieve but would not be completely unprecedented. Rather than organizing in formal unions, content moderators could also change their working conditions through worker-driven social responsibility.

1. Transnational Unions

Professors Alan Hyde and Mona Ressaissi outline four different forms of transnational worker organization.³²⁵ Unions can create formal institutions that link national union federations, national unions, or national works councils.³²⁶ Unions in different countries can partner transnationally in ad hoc campaigns.³²⁷ Unions or workers councils can enter into global framework agreements with multinational corporations.³²⁸ Additionally, networks involving actors that are not unions, such as activists or religious, women's, or indigenous groups, can develop.³²⁹

Global union federations have had some success. Hyde and Ressaissi discuss the International Transport Workers' Federation and the various European Works Councils.³³⁰ The International Transport Workers' Federation represents merchant seamen on flag-of-convenience ships and has managed to maintain internal consistency in fighting for wages.³³¹ The European Works Councils have trended toward becoming multinational works councils in the automotive industry.³³² The councils are organized by company and share information across European countries where employees are located.³³³ In 2006, the European Works Council for General Motors

Trump Administration Signals Formal Withdrawal from W.H.O., N.Y. TIMES (Oct. 22, 2020), <https://www.nytimes.com/2020/07/07/us/politics/coronavirus-trump-who.html> [<https://perma.cc/B36B-GWRG>]; Michael D. Shear, *Trump Will Withdraw U.S. from Paris Climate Agreement*, N.Y. TIMES (June 1, 2017), <https://www.nytimes.com/2017/06/01/climate/trump-paris-climate-agreement.html> [<https://perma.cc/XR2Y-CW9H>].

324. See ROBERTS, *supra* note 1, at 39.

325. See Alan Hyde & Mona Ressaissi, *Unions Without Borders: Recent Developments in the Theory, Practice and Law of Transnational Unionism*, 14 CAN. LAB. & EMP. L.J. 47, 54–55 (2008).

326. See *id.*

327. See *id.*

328. See *id.* at 55.

329. See *id.*

330. See *id.* at 58.

331. See *id.* at 57–58.

332. See *id.* at 59.

333. See *id.* at 59–60.

organized coordinated work stoppages to protest plant closings, and 40,000 workers across Europe participated.³³⁴ The IndustriALL Global Union represents fifty million mining, energy, and manufacturing workers in 140 countries.³³⁵ IndustriALL brought together affiliates of the former global union federations of the International Metalworkers' Federation; the International Federation of Chemical, Energy, Mine and General Workers' Unions; and the International Textile, Garment and Leather Workers' Federation.³³⁶ The global union federation has negotiated global framework agreements with over fifty multinational corporations.³³⁷

There are no unions dedicated solely to content moderators. Therefore, content moderators face numerous challenges when trying to find a place within a global union federation. Further, content moderators might find organizing under these frameworks difficult when they are separated across the globe and work for different companies and when workers do not stay in the industry for long.

Although no union is dedicated solely to content moderators in any country, examples of ad hoc campaigns of solidarity could provide guidance for the future if workers in different countries organize. The United Steelworkers attempted to partner with Japanese unions for their strike against tire manufacturer Bridgestone in the late 1990s.³³⁸ Hyde and Ressaissi note that, although the outcome was successful for the union, the process was rife with cultural misunderstandings and communication issues.³³⁹ The unions had difficulty squaring their different structures and approaches, with the U.S. union preferring demonstrations and some Japanese unions preferring peaceful consultation.³⁴⁰ This problem of cultural differences in ad hoc campaigns is also illustrated in the union recognition dispute between the UNITE HERE garment workers' union and H&M.³⁴¹ Swedish unions who sought to partner with UNITE HERE in the United States found it difficult to comprehend that a union would not be recognized and were not accustomed to public displays of controversy before any negotiations had taken place.³⁴² The UNITE campaign, which sought

334. *See id.*

335. *See Who We Are*, INDUSTRIALL GLOB. UNION, <http://www.industrialall-union.org/who-we-are> [<https://perma.cc/67QV-8L7J>] (last visited Oct. 29, 2021).

336. *See id.*

337. *See Global Framework Agreements*, INDUSTRIALL GLOB. UNION, <http://www.industrialall-union.org/global-framework-agreements> [<https://perma.cc/MSSV-G3T9>] (last visited Oct. 29, 2021).

338. *See* Hyde & Ressaissi, *supra* note 325, at 62.

339. *See id.* at 62–65.

340. *See id.*

341. *See id.* at 72.

342. *See id.*; *see also* Owen E. Herrnstadt, *Corporate Social Responsibility, International Framework Agreements and Changing Corporate Behavior in the Global Workplace*, 3 AM. U. LAB. & EMP. L.F. 263, 266 (2013) (discussing the fact that U.S. workers do not have health care, retirement security, job security, and benefits protections enjoyed by European workers, which leads U.S. unions to prioritize these issues over global framework agreements in discussions with employers).

union recognition by H&M by way of card check, achieved very little.³⁴³ The U.S. warehouse at issue was not organized by card check, as was demanded, or by a National Labor Relations Board election.³⁴⁴

Global framework agreements are agreements between multinational corporations and global union federations.³⁴⁵ Corporations consent to follow standards on fundamental labor rights, working conditions, health and safety conditions, and training in more than one country or worldwide.³⁴⁶ Global framework agreements typically recognize all of the ILO's core labor standards, and they make unions a part of enforcement and monitoring.³⁴⁷ Positive results include increased unionization among subsidiaries and increased attention to supplier working conditions.³⁴⁸

Negotiating a framework agreement between a global company and global workers is no easy feat. Successful negotiation requires both a powerful union in the company's home country and a global network of unions.³⁴⁹ As noted already, content moderators currently have neither. As of 2015, only one U.S.-based company was a signatory to a global framework agreement.³⁵⁰ A transnational union of content moderators would face the challenges of negotiating with major players based in the United States, a country where labor unions have been on the decline and where the National Labor Relations Act³⁵¹ (NLRA) exempts a large sector of the moderation workforce—*independent contractors*.³⁵²

For content moderators directly employed by large technology companies or outsourcing firms, union organization is not even in its early stages. Moreover, the NLRA prohibits "secondary boycotts" or work stoppages in support of another union by employees of an employer not directly involved in the primary labor dispute.³⁵³ Secondary boycott prohibition has been interpreted to apply when workers from one separately administered arm of a company stop work in solidarity with workers from another separately

343. See Hyde & Ressaissi, *supra* note 325, at 72.

344. See *id.*

345. See Felix Hadwiger, *Global Framework Agreements: Achieving Decent Work in Global Supply Chains?*, 7 INT'L J. LAB. RSCH. 75, 77 (2015).

346. See *id.*

347. See Hyde & Ressaissi, *supra* note 325, at 78.

348. See Hadwiger, *supra* note 345, at 88–89.

349. See Herrnstadt, *supra* note 342, at 26.

350. See Hadwiger, *supra* note 345, at 79.

351. 29 U.S.C. §§ 151–169.

352. See MICHAEL FICHTER & DIMITRIS STEVIS, FRIEDRICH EBERT STIFTUNG, GLOBAL FRAMEWORK AGREEMENTS IN A UNION-HOSTILE ENVIRONMENT: THE CASE OF THE USA (2013) (discussing the hurdles to forming global framework agreements with U.S.-based companies and noting that the state of labor relations in the United States drives down labor standards globally, while right-to-work states simultaneously drive down standards within the United States); see also *supra* Part II.C.3.c (discussing the United States's reluctance to ratify ILO conventions).

353. See 1 N. PETER LAREAU LABOR AND EMPLOYMENT LAW § 19.02 (2020).

administered division.³⁵⁴ Under the call center model, content moderators work for an outsourced company.³⁵⁵ Thus, the NLRA would prohibit call center employees from using strikes to exert pressure on the client company for which they are moderating.

However, publicity activities other than picketing are permitted under the NLRA.³⁵⁶ This proviso has been construed to allow publicity, such as handbills, for the purpose of truthfully advising the public, including consumers and members of labor organizations, that a product or products are produced by an employer with whom the labor organization has a dispute.³⁵⁷ Relatedly, content moderators could lawfully engage in corporate campaigns that seek concessions from employers by targeting upper management, consumers, suppliers, and investors with publicity and other forms of pressure.³⁵⁸

Hyde and Ressaissi note that ad hoc campaigns of transnational organizations may be unsuccessful due to a lack of general sympathy from the public.³⁵⁹ As society becomes increasingly aware of how essential content moderation is to our life online, content moderators' demands may garner greater public support. The average person may have an easier time connecting the graphic stories of content moderators encountering violence and child sexual exploitation with mental health issues than they do the importance of card-check recognition and National Labor Relations Board elections. Disruption in content moderation work would not stop a manufacturing line or close a store. The idea of a content moderator strike, on one particular platform or many, raises a scary proposition: if moderators stop working, what would the internet look like?

2. Worker-Driven Social Responsibility

Given the barriers to traditional unionization, worker organization for content moderators may come in the form of worker-driven social responsibility campaigns. The term “worker-driven social responsibility” was coined in the wake of the formation of the Coalition of Immokalee Workers (CIW), which is a human rights organization comprised largely of temporary farmworkers in Florida’s \$650 million tomato industry.³⁶⁰ In 1993, farmworkers first started meeting in Immokalee to address the abuses

354. *See generally* 29 U.S.C. § 158; *Am. Fed. of Television and Radio Artists, Wash.-Balt. Loc. v. NLRB*, 462 F.2d 887 (D.C. Cir. 1972) (holding that union members from a television station could not picket a newspaper owned by the same conglomerate).

355. *See supra* note 39 and accompanying text.

356. *See* 29 U.S.C. § 158(b)(4).

357. *See id.*

358. *See* Cynthia L. Estlund, *The Ossification of American Labor Law*, 102 COLUM. L. REV. 1527, 1605 (2002).

359. *See* Hyde & Ressaissi, *supra* note 325, at 74.

360. *See* Greg Asbed & Steve Hitov, *Preventing Forced Labor in Corporate Supply Chains: The Fair Food Program and Worker-Driven Social Responsibility*, 52 WAKE FOREST L. REV. 497, 498 (2017).

they faced in the fields, the most horrific of which was forced labor.³⁶¹ They developed an organization based on community education principles and emphasized participatory leadership with techniques designed to encourage worker participation.³⁶²

In 2005, after a widespread campaign and boycott, the CIW convinced Taco Bell to sign a “Fair Food Agreement.”³⁶³ This agreement was the first of many with companies that joined the “Fair Food Program,” which now includes Whole Foods, Trader Joes, Sodexo, McDonald’s, and Burger King.³⁶⁴ Members in the Fair Food Program pay a premium on every pound of produce they purchase from participating growers.³⁶⁵ Those premiums are then passed down to qualifying workers as a bonus.³⁶⁶ Additionally, member companies agree that they will only purchase produce from growers who are in good standing with the program, as determined by the program’s monitoring organization.³⁶⁷ The Fair Food Program also includes a human rights-based Fair Food Code of Conduct and Guidance Manual.³⁶⁸ Workers play a central role in both the formulation of and revisions to the manual.³⁶⁹ CIW teams hold worker education sessions on all Fair Food Program farms and outline workers’ rights and responsibilities.³⁷⁰ These worker education sessions arm every worker with knowledge and reference materials about their rights under the code and essentially empower the tens of thousands of workers as program monitors.³⁷¹ External program auditors conduct inspections and always interview more than half of a grower’s workforce.³⁷² These same auditors staff a 24/7 hotline where workers can file complaints.³⁷³

While the CIW’s shining achievement is the eradication of forced labor among the growers, the organization represents a model for combining consumer engagement and worker empowerment to create a program that is more than a set of idealistic goals and that truly holds large corporations and employers accountable. The CIW’s program fills the gap where the ILO conventions and corporate social responsibility agreements have previously fallen short due to insufficient worker participation and monitoring.³⁷⁴

361. *See id.* at 504.

362. *See id.* at 504–05.

363. *See id.* at 508.

364. *See id.*

365. *See id.* at 511.

366. *See id.*

367. *See id.* at 512.

368. *See id.* at 514.

369. *See id.*

370. *See id.* at 519.

371. *See id.* at 520.

372. *See id.* at 524.

373. *See id.* at 523.

374. *See* James J. Brudney, *Decent Labour Standards in Corporate Supply Chains: The Immokalee Workers Model*, in *TEMPORARY LABOUR MIGRATION IN THE GLOBAL ERA: THE REGULATORY CHALLENGES* 351, 359 (Joanna Howe & Rosemary Owens eds., 2016).

The CIW model exemplifies the potential for success in worker organization by transforming what workers themselves regard as core issues into an enforceable code and holding large corporations to that code. The CIW prioritized wage premiums, sex discrimination, and forced labor. Content moderators might prioritize health care, hazard pay, and workplace wellness standards.³⁷⁵ Call center moderators, in particular, are in a similar position to the CIW workers. The Fair Food Program threatens growers with the prospect of no longer receiving business from large multinational corporations.³⁷⁶ If the model was applied to content moderation, the code could put pressure on large companies to take business away from contractors who do not comply with the code's provisions while also compelling those companies to implement the same protections for in-house moderators. Moderators could also consider using a similar model to put pressure on major advertisers to boycott platforms that are not holding up their end of the bargain.

The workers in Immokalee benefitted from a consistent location and from years of continued organization, something content moderators lack.³⁷⁷ If a large platform or contractor were to get wind of worker organization efforts among moderators, they could simply relocate the work to another city, state, or country. This reality, along with short worker tenure, chills organization among content moderators.

However, the potential benefits of the CIW model's successes in independent auditing and worker education could greatly benefit content moderators. Moderators may feel more comfortable speaking with an auditor that employs worker advocates, some of whom may even be former moderators, than they would a counselor selected and paid for by their employer. Moderators might feel an increased sense of solidarity and empowerment by working to educate each other on the dangers of their job. Smaller community-based CIW activities like holding women's meetings, movie nights, and operating a radio station may not necessarily apply to content moderators who are geographically dispersed, but similar approaches could help content moderators feel less isolated, perhaps through the development of webinars or web-based support groups for workers who review particular kinds of content.³⁷⁸ A worker-based social responsibility system could provide the key to holding large platforms accountable for the health and well-being of their outsourced workers.³⁷⁹

375. *See supra* Part II.B.3.

376. *See* Asbed & Hitov, *supra* note 360, at 508.

377. *See* Brudney, *supra* note 374, at 352.

378. *See id.*

379. *See* FOCUS ON LAB. EXPLOITATION, WORKER-DRIVEN SOCIAL RESPONSIBILITY: EXPLORING A NEW MODEL FOR TACKLING LABOUR ABUSE IN SUPPLY CHAINS (2020), <https://www.labourexploitation.org/publications/worker-driven-social-responsibility-exploring-new-model-tackling-labour-abuse-supply> [<https://perma.cc/F4B6-4V4P>].

3. Current Worker Organization Efforts

In the Philippines, there are current attempts at worker organization in the technology field and in the call center industry, which broadly include content moderators. The BPO Industry Employees Network (BIEN) promotes the rights and welfare of BPO industry workers in the Philippines.³⁸⁰ BIEN advocates for legislation protecting BPO workers and conducts campaigns targeted at particular issues, including worker protections during the COVID-19 pandemic.³⁸¹ While the organization does not solely address issues in content moderation work, its efforts for improved working conditions, health care, and other worker protections would have a positive impact on those who perform content moderation in BPO settings in the Philippines.

In the United States, the Silicon Valley Rising campaign arose in response to the general trend of outsourcing support workers for cleaning crews, transportation, and back-office services in the technology industry.³⁸² The campaign is a coordinated campaign of labor, faith leaders, community organizations, and workers, which is fighting for better wages and a voice for Silicon Valley contractors.³⁸³ The campaign's efforts seek to research working conditions for contract workers in technology and provide adequate wages and safe and dignified working conditions for service contractors on Silicon Valley campuses.³⁸⁴ The campaign is also putting pressure on Google to negotiate a Community Benefits Agreement that will address the impact of its proposed 20,000 employee campus in San Jose.³⁸⁵ While Silicon Valley Rising is doing important work, its campaigns do not focus on content moderators specifically.

The Tech Workers Coalition is a coalition of workers “in and around the tech industry” and includes broad membership across different professions.³⁸⁶ While content moderators are welcome to join, and the coalition would likely support their efforts to organize, the organization does not specifically address their grievances.

BIEN and broader campaigns and coalitions in the United States demonstrate that worker organization in the content moderation sphere is possible. These organizations could conduct content moderator-focused

380. *See generally* BPO INDUS. EMP. NETWORK, <https://bienphilippines.wordpress.com/> [https://perma.cc/94EA-BMU9] (last visited Oct. 29, 2021).

381. *See id.*; *BIEN Sounds Alarm over COVID19 Outbreaks in Many BPOs*, BPO INDUS. EMP. NETWORK (Aug. 13, 2020), <https://bienphilippines.wordpress.com/2020/08/13/bien-sounds-alarm-over-covid19-outbreaks-in-many-bpos/#more-695> [https://perma.cc/CU82-XJQR].

382. *See About Us*, SILICON VALLEY RISING, <https://siliconvalleyrising.org/about/> [https://perma.cc/XP3P-9BEC] (last visited Oct. 29, 2021).

383. *See id.*

384. *See Responsible Contracting*, SILICON VALLEY RISING, <https://siliconvalleyrising.org/responsible-contracting/> [https://perma.cc/Q6JU-LEA6] (last visited Oct. 29, 2021).

385. *See San Jose Google*, SILICON VALLEY RISING, <https://siliconvalleyrising.org/google/> [https://perma.cc/S5T3-JN2J] (last visited Oct. 29, 2021).

386. *See generally* TECH. WORKERS COAL., <https://techworkerscoalition.org/> [https://perma.cc/8GDT-P6NM] (last visited Oct. 29, 2021).

campaigns, lend support, or serve as a general model for future organization among content moderators. When considering the suitability of these organizations to support content moderators, it is important to note the importance of the work of moderation to the technology industry. Content moderators deal directly with the platforms and apply complicated guidelines, and their decisions have broad societal impact. They may need their own dedicated organizations that emphasize the particular importance of their work and the unique workplace hazards involved in what they do.

III. CREATING “COMMUNITY GUIDELINES” THROUGH CORPORATE AND WORKER-DRIVEN SOCIAL RESPONSIBILITY

Tech workplaces have been romanticized, glamorized, and lampooned in the media. HBO’s *Silicon Valley* or the 2013 film *The Internship* depict companies with an answer for every social ill where employees sleep in nap pods and sample international cuisines free of charge.³⁸⁷ The cultural fixation on and stereotyping of the tech workplace is no accident. Technology companies actively promote their perk-filled campuses and inclusive culture.³⁸⁸ This corporate culture matches the rhetoric that the platforms themselves are communal sharing spaces where positive interactions take place.³⁸⁹

Much like the content these companies take great pains to shield us from, the working conditions of content moderators are a stark foil to the idyllic image of Silicon Valley the companies promote and we see depicted in popular culture.³⁹⁰ The internet is not the happy community it is marketed as, and neither is working for its largest purveyors. Many refer to social media as a user’s “highlight reel.” In some ways, most of the internet is a highlight reel. Content moderators bear the psychological burden of editing out the horror.

Technology companies cannot maintain the status quo when it comes to the work of content moderation.³⁹¹ Neither the companies nor the moderators seem likely to sustain current arrangements, with the exception of workers at a few smaller companies.³⁹² If nothing changes, companies will continue to face domestic and international litigation, along with public reproach.³⁹³ Journalists like Casey Newton will continue to expose horrific working conditions.³⁹⁴ Settlements and scattershot wellness programs do not protect future workers or foreign workers.³⁹⁵ As moderators continue to

387. *Silicon Valley* (HBO television broadcast); *THE INTERNSHIP* (Twentieth Century Fox 2013).

388. *See supra* note 202 and accompanying text.

389. *See supra* notes 3–4 and accompanying text.

390. *See supra* Part I.C.

391. *See supra* Part II.A.

392. *See supra* Part II.B.1.

393. *See supra* Part I.D.

394. *See supra* Part I.C.

395. *See supra* Part II.A.2.

suffer, a potential public relations disaster threatens the companies in a sector that is already the subject of increased scrutiny.³⁹⁶ The NFL concussion litigation demonstrates both the breadth of potential liability and the possibility of robust health care protections.³⁹⁷ The settlement was a victory for athletes who are suffering from head injuries.³⁹⁸ Silicon Valley has an opportunity to learn from the NFL and address content moderators' mental health now, rather than letting the issue fester and play out in the courts.

Western countries are hungry to regulate Big Tech.³⁹⁹ Legislators are grappling with moderation issues, free speech concerns, antitrust actions, and the possibility of breaking up conglomerates.⁴⁰⁰ This will undoubtedly affect content moderators.⁴⁰¹ Unfortunately, proposed and current legislation does not directly address their plight.⁴⁰² Legislation protecting content moderators or specifically regulating their work seems unlikely given the large-scale issues governments seek to tackle in the technology sector. The proposed BPO worker legislation in the Philippines is promising.⁴⁰³ The Philippines, however, recognizes that outsourcing in the technology and business process management sector generates \$25 billion in revenue, provides direct employment to 1.3 million Filipinos, and indirectly supports 3.2 million jobs.⁴⁰⁴ Trends point to legislation in various countries that will only serve to increase the scope and scale of content moderation while providing no worker protections.⁴⁰⁵

Content moderators face high barriers to unionization and organization.⁴⁰⁶ They are geographically dispersed, speak different languages, work for different companies or contractors, and often serve short stints.⁴⁰⁷ The companies themselves acknowledge that it is likely unhealthy for someone to have a "career" in content moderation.⁴⁰⁸ The structure of unions that serve a particular set of workers working for a particular company is probably not well-suited for content moderation. Some moderators are working in-house in Silicon Valley, others are working for contractors or boutique firms, and many find work piecemeal through platforms like Amazon Mechanical Turk.⁴⁰⁹ These workers have distinct needs, resulting from being

396. *See supra* Part II.A.2.

397. *See supra* Part II.A.1.

398. *See supra* note 139 and accompanying text.

399. *See supra* Part II.C.3.a.

400. *See supra* Part II.C.

401. *See supra* Part II.C.

402. *See supra* Part II.C.3.a.

403. *See supra* Part II.C.3.b.

404. *See* An Act Ensuring the Welfare and Protection of Business Process Outsourcing (BPO) Workers in the Philippines, House Bill No. 0274 (July 1, 2019) (Phil.) (citing data as of 2016).

405. *See supra* Parts II.C.1., II.C.3.

406. *See supra* Part II.D.

407. *See supra* Part I.B.

408. *See supra* note 194 and accompanying text.

409. *See supra* Part I.B.

engaged in distinct structural settings and with distinct expectations; thus, a traditional union may not be the best vehicle for addressing them.

Given the unlikelihood of government intervention or traditional forms of worker organization, how can the technology industry solve its “billion dollar problem”⁴¹⁰ and protect the health, safety, and dignity of the content moderators who are essential to its business? The answer may lie in some combination of corporate social responsibility and worker-driven social responsibility.

Social media companies market themselves as innovators who make daily life better for everyone around the globe, but they are ignoring the workers at their own doorstep. They claim time and time again in promotional videos, press releases, product releases, and marketing materials that they are trying to help solve the world’s problems.⁴¹¹ This innovative spirit should be applied to the issues facing content moderators. There may be great institutional value to bringing content moderators into the fold as technology companies come to grips with the darker reality of what is branded as a utopian and communal online (and office) space.⁴¹² Companies are the only actors currently in a position to provide moderators with the workplace and mental health support they need.⁴¹³ The work of journalists and academics, like Professor Sarah Roberts’s exposing the hidden work of content moderators, and the public’s increasing awareness of these issues put pressure on companies to act. As Professor Paul Barrett suggested, though it may be rote and repetitive, there is no reason why moderation work should be relegated to second-class status.⁴¹⁴ The companies must recognize that while their origins were in engineering and software, they now deal in content. Without moderators, their business would not exist. While there are many pitfalls to the corporate social responsibility model,⁴¹⁵ Silicon Valley has made itself uniquely publicly accountable. If technology companies devote the same resources to content moderation workers as they do to sustainability efforts or other corporate social responsibility schemes, they could solve a problem of their own making and also live up to their lofty promises.

While companies can and should institute change independently, the economic incentives to continue with a system predominantly based on outsourcing are too high for change to happen overnight.⁴¹⁶ Companies face

410. See ROBERTS, *supra* note 1, at 206.

411. See *supra* notes 3–4 and accompanying text.

412. See Newton, *supra* note 11 (noting that as companies struggle with a variety of different issues surrounding moderation, there are no greater experts in the moderation policies than those who actually moderate).

413. See *supra* Part II.D.1 (discussing the current lack of content moderator workers’ unions).

414. See *supra* note 225 and accompanying text.

415. See *supra* Part II.B.

416. See *supra* Part II.A.2.

legitimate needs for international workers that go beyond cost-cutting.⁴¹⁷ If they are to combat the mass of violative content posted daily, they need a multilingual army of moderators available 24/7.⁴¹⁸ Worker coalitions must put pressure on companies to take steps toward practices that benefit moderators. The Coalition of Immokalee Workers model may not apply perfectly to content moderators, but its spirit and philosophy could prove helpful.⁴¹⁹ A code that empowers workers, holds both contractors and corporations accountable, and addresses priority workplace issues articulated by those workers could prove to be the solution to the issues inherent in the call center model,⁴²⁰ a model some in the industry acknowledge is unsustainable.⁴²¹ Content moderators are not as unified as farmworkers who work in one general area, but their access to technology could help in cross-border organization.⁴²² Broad campaigns that raise general user awareness could also put pressure on companies to adopt stronger corporate social responsibility measures of their own.

The mind of a philosopher. The gut of a police detective. The heart of a kindergarten teacher. We both need and expect content moderators to have these things. To wear these different hats, content moderators must be healthy, and they cannot achieve adequate mental health without adequate support. Through a combination of corporate and worker-driven social responsibility, content moderators may be able to pressure the industry to formulate their own “community guidelines,” bring their work out of the shadows, and guarantee they get the support that they deserve.

CONCLUSION

Whether they are referred to as “community guidelines,” “community standards,” or “rules and policies,” technology companies have created frameworks that turn online platforms into spaces that users and advertisers want to engage with. Content moderators who implement these guidelines on a daily basis do not have guidelines of their own to navigate their high-stress jobs. Low pay, lack of health care, and poor working conditions will continue to plague content moderators until such guidelines are established.

Moderators cannot rely solely on litigation or governmental regulation for relief. Worker-driven social responsibility campaigns can help them standardize current corporate social responsibility efforts on moderator wellness and hold companies accountable. The worker-driven social responsibility model would allow workers to set their own priorities, such as hazard pay, better physical and mental health care, and workplace wellness requirements. Worker-driven campaigns can bridge the gaps that exist in

417. *See supra* Part I.B.

418. *See supra* Part I.A.

419. *See supra* Part II.D.2.

420. *See supra* Part II.D.2.

421. *See supra* note 194 and accompanying text.

422. *See supra* Part II.D.3.

current corporate efforts and ensure that the “community guidelines” for content moderators are as comprehensive as the platform guidelines they enforce.